AD-A069 780 ILLINOIS UNIV AT URBANA-CHAMPAIGN COORDINATED SCIENCE LAB F/G 9/3
ON OPTIMUM DATA QUANTIZATION FOR SIGNAL DETECTION.(U)
SEP 78 D ALEXANDROU
UNCLASSIFIED R-827 DAAB07-72-C-0259 NL

| OF |
AD
A069780

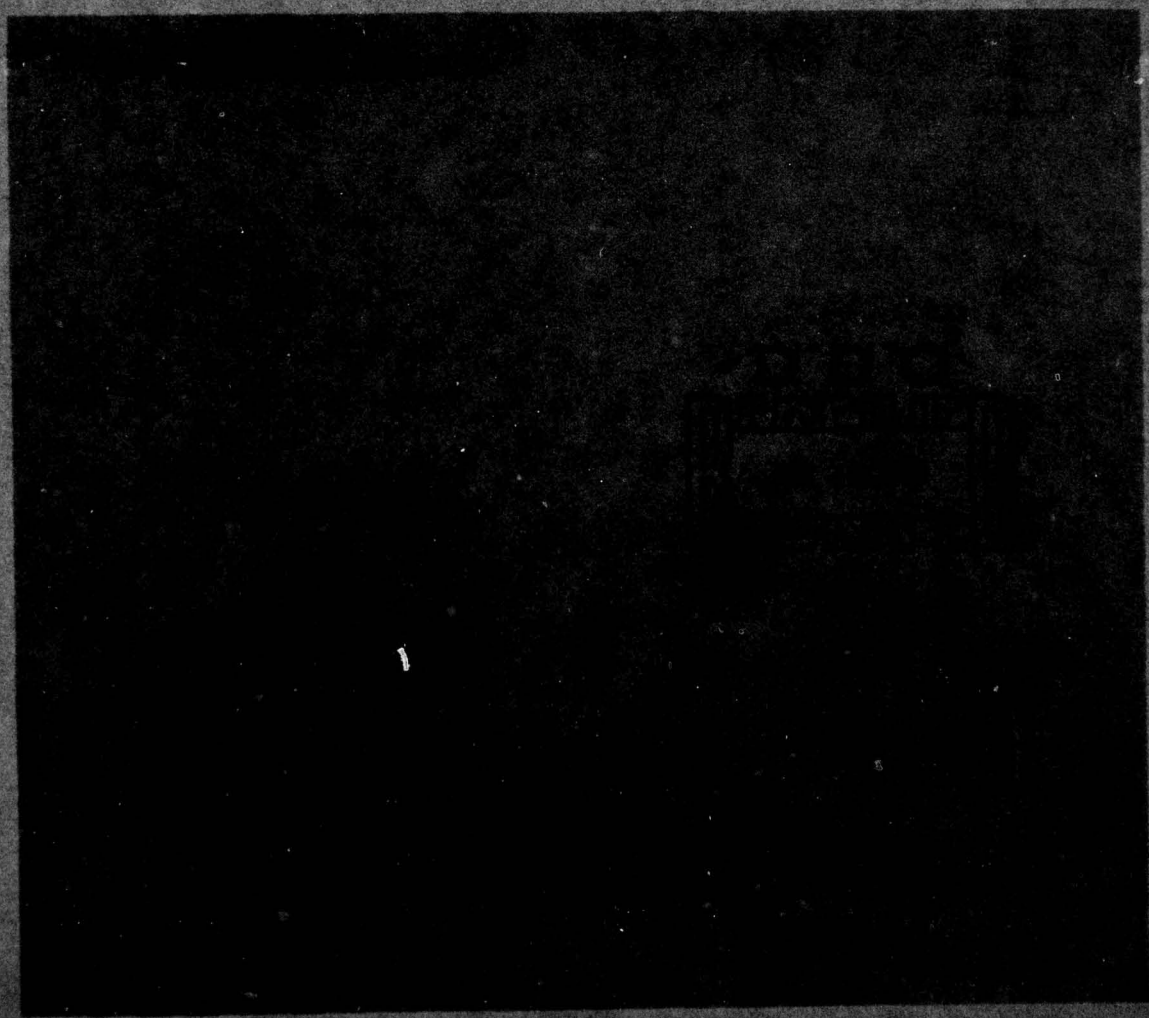
END
DATE
FILMED
7-79
DDC

REPORT R-827 SEPTEMBER, 1978 UILU-ENG 78-2210

CSL COORDINATED SCIENCE LABORATORY

LEVEL II

ON OPTIMUM DATA QUANTIZATION FOR SIGNAL DETECTION



UNIVERSITY OF ILLINOIS - URBANA, ILLINOIS

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER

2. GOVT ACCESSION NO.

3. RECIPIENT'S CATALOG NUMBER

(9) Master's thesis

4. TITLE (and Subtitle)

(4) ON OPTIMUM DATA QUANTIZATION FOR SIGNAL DETECTION

5. TYPE OF REPORT & PERIOD COVERED

(9) Technical Report

6. PERFORMING ORG. REPORT NUMBER

(14) R-827; UILU-ENG-78-2220

7. AUTHOR(s)

(10) Dimitrios Alexandrou

8. CONTRACT OR GRANT NUMBER(s)

(15) DAAB-07-72-C-0259

9. PERFORMING ORGANIZATION NAME AND ADDRESS

Coordinated Science Laboratory
University of Illinois at Urbana-Champaign
Urbana, Illinois 61801

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS

11. CONTROLLING OFFICE NAME AND ADDRESS

Joint Services Electronics Program

12. REPORT DATE

(11) September, 1978

13. NUMBER OF PAGES

58

14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)

(12) 66 p.

15. SECURITY CLASS. (of this report)

UNCLASSIFIED

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Quantization
Signal Detection

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

An introduction to quantization and to several important detection problems is given in the initial sections. A detailed review follows of most of the work done on quantization for detection. The equivalence of the criterion of minimum mean-squared error between quantized data and data transformed by the locally-optimum nonlinearity and the one of maximum efficacy is shown for the general case of local decisions based on independent samples. In addition, a sufficient condition for optimum detection is derived for the above case. Finally, numerical results are obtained for the locally-optimum quantizer for the case

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

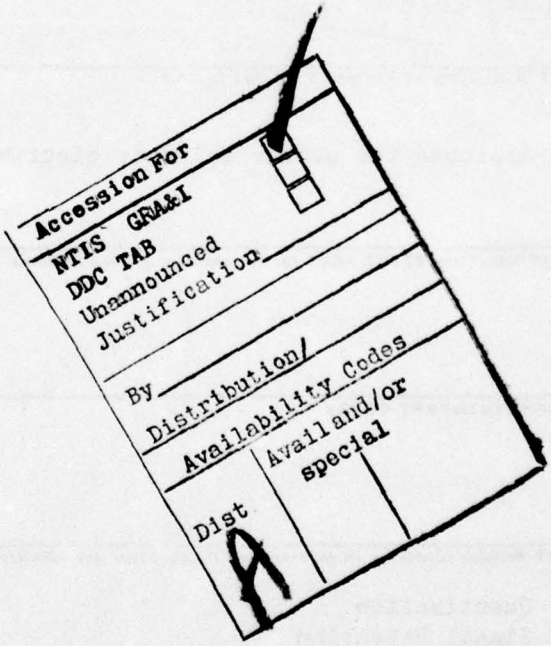
097 700

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

20. ABSTRACT (continued)

of detecting stochastic signals in generalized Gaussian noise.



UILU-ENG 78-2220

ON OPTIMUM DATA QUANTIZATION FOR SIGNAL DETECTION

by

Dimitrios Alexandrou

This work was supported in part by the Joint Services Electronics Program (U.S. Army, U.S. Navy and U.S. Air Force) under Contract DAAB-07-72-C-0259.

ON OPTIMUM DATA QUANTIZATION FOR SIGNAL DETECTION

BY

DIMITRIOS ALEXANDROU

B.S., University of Illinois, 1977

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Master of Science in Electrical Engineering
in the Graduate College of the
University of Illinois at Urbana-Champaign, 1978

Thesis Advisor: Professor H. V. Poor

Urbana, Illinois

ABSTRACT

An introduction to quantization and to several important detection problems is given in the initial sections. A detailed review follows of most of the work done on quantization for detection. The equivalence of the criterion of minimum mean-squared error between quantized data and data transformed by the locally-optimum nonlinearity and the one of maximum efficacy is shown for the general case of local decisions based on independent samples. In addition, a sufficient condition for optimum detection is derived for the above case. Finally, numerical results are obtained for the locally-optimum quantizer for the case of detecting stochastic signals in generalized Gaussian noise (both additive-noise and scale-change model.)

ACKNOWLEDGMENTS

I would like to sincerely thank Professor H. V. Poor for his help and guidance.

TABLE OF CONTENTS

INTRODUCTION

Despite the development of new coding schemes such as permutation codes, tree codes etc., simple quantization continues to be a very popular method of analog-to-digital conversion. The conceptual simplicity of the quantizers, their near optimum performance and the fact that they can be readily implemented in hardware are the main reasons for their popularity. Because of their diversified use, quantizers have been optimized based on several different criteria. Before we attempt to give an overview of the work done in the area of optimum quantization, we will first describe the basic quantizer equations.

A quantizer Q with M levels can be represented as a pair $(\vec{t},\vec{q})$ where $\vec{q} \in \mathbb{R}^M$ are the levels of Q; the breakpoints $\vec{t} \in \mathbb{R}^{M+1}$ are such that $-\infty = t_0 < t_1 < \ldots < t_{M-1} < t_M = \infty$. We take $Q(x) = q_k$ when $x \in (t_{k-1}, t_k]$ for $k = 1,\ldots,M$. Let X be a scalar random variable with probability density f(x). Two widely accepted criteria in terms of which the performance of the quantizer is defined are the distortion

$$D = \sum_{k=1}^{M} \int_{t_{k-1}}^{t_k} g(x-q_k) f(x) dx$$

and the entropy

$$H(Q) = -\sum_{k=1}^{M} (\log_2 f_k)\, f_k$$

where g is a non-negative weighting function and

$$f_k = \int_{t_{k-1}}^{t_k} f(x) dx.$$

In the early literature, the parameters of an "optimum" quantizer were chosen to optimize D or H or a combination of the two.

Max [1] first considered the problem of designing an optimum quantizer with minimum distortion as the criterion of performance. Note that for $g(x) = x^2$ the distortion function D becomes the mean-squared-error expression between the input and the output of the quantizer:

$$D = \sum_{k=1}^{M} \int_{t_{k-1}}^{t_k} (x-q_k)^2 f(x)dx.$$

When the criterion is minimum mean-squared error, Max showed that the parameters of the optimum quantizer satisfy the following equations:

$$q_k = \frac{\int_{t_{k-1}}^{t_k} xf(x)dx}{\int_{t_{k-1}}^{t_k} f(x)dx} \; ; \; k = 1,\ldots,M \qquad (1)$$

$$t_k = \tfrac{1}{2} (q_k + q_{k+1}) \; ; \; k = 1,\ldots.M-1. \qquad (2)$$

The analytical solution of these equations is impossible for all but trivial cases. A numerical solution, however, is straightforward. Many iteration techniques are feasible and one is given by Max.

In addition, Roe [2] has proposed an approximation, based on Max's equations which is of practical interest and yields near-optimum results. Further simplification of the structure of the optimum quantizer results from Algasi's [3] work. First, he derived approximate expressions for the distortion for the case of uniform* quantization. Then, by deriving similar approximations for a non-uniform quantizer he concluded that, depending on the number of quantization levels, a uniform quantizer may perform equivalently to a non-uniform one.

*Equally spaced breakpoint and output levels.

Although minimum distortion is desirable, it is not the absolute criterion for the performance of a quantizer. Several authors [4] have indicated that the entropy H(Q) of the output of MMSE quantizer is high. Since H, in general, is the minimum amount of information which must be transmitted in order to achieve arbitrarily small probability of erroneous detection, high H(Q) is undesirable. It was shown by Messerschmitt [5] that, for certain input distributions, the minimum-distortion quantizer and the maximum-output-entropy quantizer are approximately the same. These results indicate that a trade-off between low distortion and high output entropy is unavoidable. One way to approach the problem would be to minimize H(Q) for a fixed value of D. Therefore, in general, a combination of D and H(Q) should be used to define the appropriate performance criterion for optimum quantization.

In the work reviewed thus far, the quantizer is treated as a simple coding scheme which is used to facilitate signal transmission. However, in the above discussion, no mention is given of how a quantizer can be used for signal detection. Since the latter is our main interest here, we will proceed with a review of the detection problem to eventually concentrate on quantization for optimum detection.

1. THE GENERAL DETECTION PROBLEM

The M-ary communication problem requires the design of a receiver that will decide, with minimum probability of error, which of M possible signals has been sent. In general, under hypothesis H_m, the receiver observes

$$Y(t) = S_m[t,\Theta(t)] + N(t) \; ; \; 0 \leq t \leq T, \quad 1 \leq m \leq M$$

where N(t) is a noise process with arbitrary statistics and $S_m[t,\Theta(t)]$ is the m^{th} signal. $\Theta(t)$ represents unknown channel effects on the signal. To simplify the analysis, assume that there is a discrete representation of the problem, obtained through time sampling. If the dimension of the discrete representation is n, the receiver observes

$$\underline{Y} = \underline{S}m(\underline{\Theta}) + \underline{N} \; ; \; m = 1,2,\ldots,M$$

where

$$\underline{Y} = (y_1,\ldots,y_n)$$

$$\underline{N} = (n_1,\ldots,n_n)$$

$$\underline{\Theta} = (\theta_1,\ldots,\theta_n)$$

$$\underline{S}_m(\underline{\Theta}) = (S_{m1}(\underline{\Theta}),\ldots,S_{mn}(\underline{\Theta})).$$

One of the M signals $\underline{S}_m(\underline{\Theta})$, $1 \leq m \leq M$ is sent. $\underline{\Theta}$ represents a set of K "nuisance" parameters (e.g. unknown amplitude and/or unknown phase). The additive noise has a multivariate density function $f_N(\underline{N})$.

A. Binary Detection - Constant Signal Case

The simplest hypothesis testing problem represents the detection of a constant signal in additive white noise. The hypothesis pair is given by

$$H_0 : y_i = n_i$$

versus $\qquad i = 1,\ldots,N \qquad (3)$

$$H_1 : y_i = \theta + n_i .$$

A decision rule δ will be of the type:

$$\delta = \begin{cases} 0 \text{ , if } H_0 \text{ is accepted} \\ 1 \text{ , if } H_1 \text{ is accepted.} \end{cases} \qquad (4)$$

We assume that there exists a conditional p.d.f. $f_N(\underline{y}|\Theta)$ where $\Theta \in \{0,\theta\}$. A prior distribution of Θ and a cost function $C(\Theta,\delta)$ may or may not exist. When they both exist, then minimum average cost can be used as the criterion for optimum detection (Bayes rule). If the prior distribution is unknown but a cost function is defined, then a Bayes rule under the worst possible prior distribution (Minimax rule) can be applied. Finally, if the prior distribution is unknown and no information is given about the cost function, the Neyman-Pearson decision rule can be used.

B. The Neyman-Pearson Criterion

The following two types of errors can occur in binary detection

Type I error: Choose H_1 when H_0 is true (false alarm)

Type II error: Choose H_0 when H_1 is true (miss).

Let:

$$P_F(\delta) = P\{\delta(\underline{y}) = 1 \mid \Theta = 0\} = \text{probability of false alarm}$$

$$P_M(\delta) = P\{\delta(\underline{y}) = 0 \mid \Theta = \theta\} = \text{probability of a miss}$$

$$P_H(\delta) = P\{\delta(\underline{y}) = 1 \mid \Theta = \theta\} = \text{probability of a hit.}$$

Also, let

$$f_N(\underline{y} \mid \Theta = 0) = f_N(\underline{y} \mid H_0)$$

$$f_N(\underline{y} \mid \Theta = \theta) = f_N(\underline{y} \mid H_1).$$

The objective of the Neyman-Pearson criterion is to constrain δ to be such that $P_F(\delta) \leq \alpha_0$, where α_0 is a prescribed bound, and then find a decision

rule $\delta = \tilde{\delta}_{NP}$ which maximizes $P_H(\delta)$ within the constraint. In general, $\tilde{\delta}_{NP}$ is given by:

$$\tilde{\delta}_{NP}(\underline{y}) = \begin{cases} 1 \ ; \ \text{if } f_N(\underline{y}|H_1) > \eta_0 f_N(\underline{y}|H_0) \\ \gamma_0; \ \text{if } f_N(\underline{y}|H_1) = \eta_0 f_N(\underline{y}|H_0) \\ 0 \ ; \ \text{if } f_N(\underline{y}|H_1) < \eta_0 f_N(\underline{y}|H_0) \end{cases} \tag{5}$$

where $\tilde{\delta}_{NP}(\underline{y})$ is the probability with which we accept H_1 when $\underline{y}$ is observed and $0 \le \gamma_0 \le 1$; γ_0 and η_0 are chosen such that $P_F(\tilde{\delta}_{NP}) = \alpha_0$.

The quantity $L(\underline{y}) = \dfrac{f_N(\underline{y}|H_1)}{f_N(\underline{y}|H_0)}$ is known as the "likelihood ratio" for testing the hypothesis pair H_0 vs H_1.

C. Detection in Non-Gaussian Noise.

The statistics of $L(\underline{y})$ are vital to optimum detection. The expression for $L(\underline{y})$ is greatly simplified if $f_N(\cdot)$ is assumed to be Gaussian. This assumption was made in most of the early literature on detection theory. In many practical cases, however, more severe types of noise are encountered. If the detector is based on Gaussian noise assumptions, performance deteriorates. If, on the other hand, non-Gaussian p.d.f.'s are assumed, determination of the statistics of $L(\underline{y})$ becomes extremely difficult.

Some simplification results from the assumption that the noise samples $\{n_i : i=1,2,\ldots,N\}$ are mutually statistically independent (i.e. time samples spaced sufficiently apart) and identically distributed with a common p.d.f. $f(\cdot)$. Then,

$$L(\underline{y}) = \prod_{i=1}^{N} \frac{f(y_i|H_1)}{f(y_i|H_0)}$$

$$\Rightarrow \log L(\underline{y}) = \sum_{i=1}^{N} \wedge_i \tag{6}$$

where

$$\wedge_i \triangleq \log \frac{f(y_i|H_1)}{f(y_i|H_0)}, \ i = 1,\ldots,N.$$

In this form, the test statistic log $L(\underline{y})$ is the sum of N independent random variables and, although it is theoretically possible to obtain its p.d.f., actual analytical evaluation is very cumbersome except for special cases.

Simulation techniques (e.g. Monte Carlo) can be used to determine the p.d.f. of log $L(\underline{y})$ for any given N and $f(\cdot)$. Another practice commonly found in literature is the assumption that if the receiver integrates a sufficiently large number of independent samples, the resulting distribution of the test statistic will be Gaussian. If the variance of the noise is large, however, the fundamental limit theorem cannot practically be invoked unless there is some noise suppresion before integration (e.g. by clipping or limiting the received signal), the usual justification being that this will increase the signal-to-noise ratio (SNR). But then, this evaluation of performance based on SNR may be of little value in indicating information rate (or reliability) of the detection scheme [6].

In general, the test statistic is simplified when the signal is very weak compared to the noise (local detection). Several authors turned their attention to the problem of detection in non-Gaussian interference under weak signal assumptions. Concentration on this problem is justified by real-life cases, where signals are frequently very weak compared to the noise. Also, it is obvious that a large signal would be easier to detect and the local detector will, in general, perform satisfactorily for larger signals as well.

D. Local Detection

As pointed out previously, "local" detection is an expression used to describe the situation where the signals are very weak compared to the

noise. The expression "threshhold" detection is also used, mostly in the early literature. Middleton [7] was the first author to consider the local detection problem and he obtained a receiver that was, essentially, a cross-correlator. Examining the same problem and using a technique similar to Middleton's, Rudnick [8] showed that the optimum receiver must be nonlinear. The discrepancy was solved by Algasi & Lerner [9] in Rudnick's favor. They showed that, for arbitrary noise, Middleton failed to include all of the necessary terms in his power series expansion of the noise p.d.f. They also discussed the problem of actual implementation of the receiver, and, more importantly, they showed that under certain conditions the optimum receiver takes a canonical form. The canonical receiver consists of a nonlinearity which depends on the additive noise, followed by a receiver which is optimum for detection in Gaussian noise. Antonov [10] derived the same receiver for a slightly more general class of signals and he examined its asymptotic performance (number of samples approaches infinity). Finally Ribin [11] showed that for infinite observation time (or infinite number of samples) the local receiver yields a probability of error no higher than any other receiver. That is, it is asymptotically optimum.

The basic procedure followed by all authors mentioned above was the following: the p.d.f. of the arbitrary noise process was expanded in a power series (Taylor series expansion) and then the small signal assumption was used to dispose of a number of terms that would, under this assumption, be insignificant. Hence, simplification is achieved.

E. Locally-Optimum Detection

A different approach to the local problem that yields, essentially, the same results is the following: the Neyman-Pearson criterion is applied,

as described before, with one modification. Instead of maximizing the power function $P_H(\delta)$, the "locally-optimum" detector maximizes the slope with respect to θ of the power function at the origin while still keeping a fixed false-alarm probability. That is, the locally-optimum detector maximizes $\frac{\partial}{\partial\theta} P_H(\delta)$ subject to $P_F(\delta) \leq \alpha_0$. It can be shown [12], that the locally-optimum test is given by

$$\Lambda(\underline{y}) \triangleq [f_0(\underline{y})]^{-1} \frac{\partial f_1(\underline{y},\theta)}{\partial\theta}\bigg|_{\theta=0} \begin{cases} > \tau \Rightarrow H_1 \\ \\ < \tau \Rightarrow H_0 \end{cases} \tag{7}$$

where

$$f_0(\cdot) = f(\cdot|H_0) \text{ and } f_1(\cdot) = f(\cdot|H_1).$$

Example - Constant Signal in Additive Noise.

As before, the hypothesis pair is given by

$$\begin{array}{ll} H_0 : y_i = n_i & \\ \text{vs} & i=1,\ldots,N. \\ H_1 : y_i = \theta + n_i & \end{array}$$

The noise samples are assumed to be a set of real-valued, mutually independent, identically-distributed random variables with a common p.d.f. $f_n(x)$. For simplicity, assume $\theta > 0$.

The optimum (Neyman-Pearson) test is given by

$$\begin{cases} \sum_{i=1}^{N} g_o(y_i) > T' \Rightarrow H_1 \\ \\ \sum_{i=1}^{N} g_o(y_i) < T' \Rightarrow H_0 \end{cases} \tag{8}$$

where

$$g_o(x) = \log[f_n(x-\theta)/f_n(x)].$$

The locally optimum test, for the case at hand, reduces to

$$\begin{cases} \sum_{i=1}^{N} g_{1o}(y_i) > T'' \Rightarrow H_1 \\ \\ \sum_{i=1}^{N} g_{1o}(y_i) < T'' \Rightarrow H_0 \end{cases} \tag{9}$$

where

$$g_{1o}(x) = \frac{\partial}{\partial\theta} \log f_n(y_i-\theta)\Big|_{\theta=0} = - \frac{f'_n(y_i)}{f_n(y_i)} .$$

It can be seen that the locally-optimum detector is, for this case and in general, considerably simpler in structure than the Neyman-Pearson optimum detector. For this reason and because of its practical importance, the locally-optimum detector has been studied extensively.

F. Asymptotic Efficiency of the Locally-Optimum Detector

The question arises as to how does the locally-optimum detector's performance compare with the performance of the strictly optimum detector. Answering that question, Capon [13] showed that the locally-optimum detector is, in some sense, as efficient asymptotically as the Neyman-Pearson optimum detector. This comparison is based on the concept of asymptotic relative efficiency (ARE).

Suppose that two detectors are designed to detect the same signal and with the same probability of correct detection. Suppose, further, that the two detectors require sample sizes n_1 and n_2 respectively to achieve the prescribed error probabilities. If $n_1<n_2$, it is intuitively justifiable to say that the first detector is more "efficient" than the other. A rigorous definition of ARE follows.

Let $N_i\ (\alpha,\beta,\theta)$ denote the number of samples that a detector D_i requires in order to achieve a false-alarm probability α, and a probability

of correct detection at least equal to β with signal strength θ. Then the asymptotic relative efficiency (ARE) of a detector D_2 with respect to a reference detector D_1 is

$$ARE_{2,1} = \lim_{\substack{\theta \to 0 \\ N_1 \to \infty \\ N_2 \to \infty}} N_1(\alpha,\beta,\theta) \; / \; N_2(\alpha,\beta,\theta) \tag{10}$$

If the detectors D_1 and D_2 are based on the statistics W_1 and W_2 respectively then Capon, based on a theorem by Pitman [14], showed that, subject to some regularity conditions

$$ARE_{1,2} = E_{W_1} \; / \; E_{W_2} \tag{11}$$

where

$$E_{W_i} = \lim_{n \to \infty} \left\{ \left[\partial E_\theta \{W_i\} / \partial\theta \big|_{\theta=0} \right]^2 / n \; Var_0(W_i) \right\} \tag{12}$$

is the <u>efficacy</u> of detector D_i.

Therefore, the test with the higher efficacy is the most efficient asymptotically. Capon showed the efficacies of the Neyman-Pearson optimum and the locally optimum detector to be equal. This means that, asymptotically, the locally optimum detector is as efficient as the strictly optimum (Neyman-Pearson) detector.

In addition, Miller and Thomas [15] examined the ARE of both the optimum and locally optimum detectors using a linear detector as a reference. The cases of a constant as well as a time varying signal were considered for several general classes of noise p.d.f.'s. They also found that the form of both g_o and g_{lo} depends in very critical ways on the exact noise density. This type of annoying dependence led to efforts by many authors to design a detector that would be insensitive to variations of the noise statistics.

G. Robust Detection

Classical detection procedures, some of which were discussed here, require exact knowledge of the statistical properties of the noise. Most often, the functional form of the noise distribution is assumed to be known, with only a finite number of unknown parameters. Much research has been concentrated on these "parametric" detection problems. However, in many practical cases of digital communication, no information is available about the form of the noise distribution. In this case, the detection problem is "non-parametric". Correspondingly, a detector designed without the knowledge of the functional form of the noise distribution is a non-parametric (or distribution-free) detector.

The main characteristic of the non-parametric detector is its robustness, i.e. it guarantees a minimum performance level over large classes of noise distributions. The robustness of the non-parametric detectors is usually measured in terms of ARE using as a reference a detector designed for Gaussian noise. A review of the more important non-parametric techniques can be found in [16].

Some of the best known non-parametric detectors are based on signs or "ranks" of the received data samples. The "rank" detectors compare very favorably to the optimum (parametric) detectors in many cases, but they are considerably harder to implement. On the other hand, detectors based only on the signs of the data are very easy to implement but they do not perform as well. Several attempts have been made to improve the performance of the sign detector while retaining much of its simplicity. Ching and Kurtz [17] proposed the "m-interval detector". It was designed on the basis of a finite set of parameters of the noise distribution

rather than its functional form. It was shown to be robust with little loss of efficiency. However, assumptions were made about some characteristics of the noise distribution.

Kassam and Thomas [18], on the other hand, derived generalizations of the sign detector which are completely non-parametric. The only assumption made was that the noise p.d.f. is symmetric. They showed that these detectors, based on the application of a conditional test, have much better detection performance than the simple sign detector, with implementation remaining relatively simple.

In general, non-parametric methods tend to be too conservative because they fail to exploit some further information, even if incomplete, that might be available about a particular class of noise statistics. In these cases, it would be of interest to design robust detectors maximizing the worst case performance over the whole class about which the information is available. One such class was considered by Kassam and Thomas [19] and the results were applied to obtain robust detector structures for contaminated nominal densities in a specific class of density functions which includes the Gaussian.

2. OPTIMUM QUANTIZATION FOR SIGNAL DETECTION

Not until very recently were quantizers considered as a possible solution to detection problems, some of which were reviewed here. Previously, most of the work on quantization had been based on mean-squared-error or entropy-based optimality criteria.

Kassam [20] first approached the quantization problem with the objective to use the quantized data to form a test of hypothesis for signal detection. He studied the design of quantizers to be used in place of the test function in the case of detection of known signals in additive noise, based on independent samples and he showed that the criteria of maximum ARE and maximum local power slope lead to the same quantizer design. Based on Kassam's results, Poor and Thomas [21] extended his work to the general problem of local decisions.

A careful analysis of both of these results will follow, since we are going to rely heavily on them.

A. Known Signal, Additive Noise Case

Let $\{X_i\}_{i=1}^{N}$ be a sequence of N independent samples, described by

$$X_i = \theta S_i + N_i \ , \ i \le i \le N \ , \ \theta \ge 0 .$$

$\{S_i\}_{i=1}^{N}$ is a known signal sequence and $\{N_i\}_{i=1}^{N}$ is a sequence of independent, identically distributed noise samples with common density and distribution functions $f(\cdot)$ and $F(\cdot)$ respectively. Also, assume that $f(\cdot)$ is symmetric about its origin and absolutely continuous. For the local case, we consider the limit as $\theta \to 0$.

The hypothesis pair to be tested is

$$H_0 : \theta = 0$$

versus

$$H_1 : \theta > 0 .$$

For the case at hand, the locally optimum test is given by

$$\begin{cases} \sum_{i=1}^{N} g_{1o}(x_i)\, S_i > \tau \Rightarrow H_1 \\ \\ \sum_{i=1}^{N} g_{1o}(x_i)\, S_i < \tau \Rightarrow H_0 \end{cases} \tag{13}$$

where, as before, $g_{1o}(\cdot) = \frac{-f'(\cdot)}{f(\cdot)}$.

In his paper, Kassam considered the following problem. Given a positive integer M, design an M-level quantizer $Q_M(\cdot)$ such that, with Y_i defined by

$$Y_i = Q_M(x_i),$$

the statistic

$$S = \sum_{i=1}^{N} Y_i S_i$$

is optimum for deciding between H_0 and H_1. In essence, the objective is to replace the locally-optimum nonlinearity $g_{1o}(\cdot)$, with the output of an M-level quantizer. The problem then is to find a quantizer that, when used in this way, will produce a test statistic that will result in maximum detection performance from among all M-level quantizers.

Kassam assumes that the optimum quantizer is symmetric and later he shows that the symmetric quantizer does indeed maximize performance from among all M-level quantizers for symmetric noise densities. He defines the symmetric 2m-level quantizer as follows:

The positive input values are partitioned into m intervals $T_1,\ldots,T_m$ where $T_j = [t_j, t_{j-1})$ and $\{t_j\}_{j=0}^{m}$ is a decreasing sequence of non-negative

numbers with $t_0 = \infty$ and $t_m = 0$. The output level corresponding to T_j is denoted by q_j. For negative input values, $T_{-j} = (-t_{j-1}, -t_j]$ and $q_{-j} = -q_j$ for $1 \le j \le m$. Therefore, the symmetric 2m-level quantizer is completely defined in terms of the parameter vectors $\underline{t} = (t_1, t_2, \ldots, t_{m-1})$ and $\underline{q} = (q_1, q_2, \ldots, q_m)$.

I. The Quantizer for Maximum Detection Efficacy

The efficacy G of a local test for H_0 vs H_1 based on test statistic S, for the constant signal, additive noise case is defined (see Eq. (12)) by

$$G = \lim_{N\to\infty} \frac{1}{N} \frac{\left[\frac{d}{d\theta} E_\theta\{S\}\big|_{\theta=0}\right]^2}{\mathrm{Var}_0\{S\}} . \tag{14}$$

As noted in a previous section, the efficacy is an asymptotic measure of performance of the local test. Using efficacy as a measure of performance, Kassam first looked for the optimum 2m-level quantizer maximizing efficacy. In terms of the quantizer parameters the efficacy becomes

$$G = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} S_i^2 \; \frac{2\left\{\sum_{j=1}^{m} q_j[f(t_j)-f(t_{j-1})]\right\}^2}{\sum_{j=1}^{m} q_j^2[F(t_{j-1})-F(t_j)]} . \tag{15}$$

Maximizing G is equivalent to maximizing E, where

$$E = \frac{2\left\{\sum_{j=1}^{m} q_j[f(t_j)-f(t_{j-1})]\right\}^2}{\sum_{j=1}^{m} q_j^2[F(t_{j-1})-F(t_j)]} \tag{16}$$

with the mild assumption that $\lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} S_i^2$ exists and is finite.

The author proceeds to maximize E with respect to $\underline{q}$ for a given vector $\underline{t}$. By taking the partial derivatives $\partial E/\partial q_j$ and setting them equal to zero, he finds that the elements of the optimum vector $\underline{q}^*$ are given by

$$q_j^* = \frac{f(t_j)-f(t_{j-1})}{F(t_{j-1})-F(t_j)} \quad , \quad j = 1,2,\ldots,m \cdot \tag{17}$$

Substituting the values defined by (17) for q_j, the normalized efficacy expression (16) becomes

$$E^* = 2 \sum_{j=1}^{m} \frac{[f(t_j)-f(t_{j-1})]^2}{[F(t_{j-1})-F(t_j)]} \cdot \tag{18}$$

E^* can now be maximized with respect to $\underline{t}$ in order to complete the specification of the optimum quantizer.

II. The Optimum Quantizer as an Approximation to the Locally-Optimum Nonlinearity

Another performance criterion considered by Kassam is the one of minimum-squared-error between the locally optimum nonlinearity $g_{lo}(x) = \frac{-f'(x)}{f(x)}$ and the output of the optimum quantizer $q_{2m}(x)$. The squared-error expression is given by

$$\epsilon = E\left\{\left[q_{2m}(x_i) + \frac{f'(x_i)}{f(x_i)}\right]^2\right\} \cdot \tag{19}$$

In term of the quantizer parameters this expression becomes

$$\epsilon = 2 \sum_{j=1}^{m} q_j^2 \int_{t_j}^{t_{j-1}} f(x)dx + 4 \sum_{j=1}^{m} q_j \int_{t_j}^{t_{j-1}} f'(x)dx + I_f \tag{20}$$

where

$$I_f = E\left\{\left[\frac{f'(x_i)}{f(x_i)}\right]^2\right\} = \int_{-\infty}^{\infty} \frac{[f'(x)]^2}{f(x)} dx \cdot$$

Obviously, ϵ is finite only if I_f is finite and this assumption is made by Kassam.

Again, by taking the partials with respect to $\underline{q}$, for a given vector $\underline{t}$ and setting them equal to zero, he finds that ϵ is minimized when the

output levels are given by

$$q_j^* = \frac{f(t_j)-f(t_{j-1})}{F(t_{j-1})-F(t_j)} \quad . \tag{21}$$

Note that (21) is identical to Equation (17). Substituting for q_j into Equation (20), the minimized value of ϵ is given by

$$\epsilon^* = I_f - 2 \sum_{j=1}^{m} \frac{[f(t_j)-f(t_{j-1})]^2}{F(t_{j-1})-F(t_j)} = \tag{22}$$

$$= I_f - E^* .$$

Equation (22) is an important result because it clearly indicates that continuing the minimization of ϵ^* with respect to $\underline{t}$ is equivalent to maximizing E^*. Therefore, Kassam concludes, the quantizer minimizing the mean-squared-error between quantized data and data transformed by the locally optimum nonlinearity is the same as the quantizer maximizing the detection efficacy. From Equation (22), setting the partial derivatives of ϵ^* with respect to $\underline{t}$ equal to zero, Kassam obtains

$$\frac{q_{j+1} + q_j}{2} = - \frac{f'(t_j)}{f(t_j)} = g_{lo}(t_j) \quad , \quad j=1,2,\ldots,(m-1). \tag{23}$$

The solutions of the set of Equations (21) and (23) give the parameter values of the optimum quantizer.

It is important to note that the conditions for optimality derived by Kassam are only necessary conditions. He does not examine the sufficiency of either (21) or (23).

B. The General Problem of Local Decisions

Poor and Thomas [21] extended Kassam's results to the general problem of local decisions. The general problem is formulated as follows.

Assume that we observe a sequence $\{x_i\}_{i=1}^{N} = \underline{x}$ of independent real samples and that we have a corresponding sequence $\{P_\theta^{(i)}; \theta \in \Theta \subseteq R\}_{i=1}^{n}$ of indexed classes of distributions on the real line. For a particular θ_0 in a right-open set of Θ we wish to test

$$H_{\theta_0} : X_i \sim P_{\theta_0}^{(i)} \; ; \; i=1,2,\ldots,n$$

vs (24)

$$H_\theta : X_i \sim P_\theta^{(i)} \; ; \; i=1,2,\ldots,n$$

where $\theta > \theta_0$. For the local case, consider the limit as $\theta \to \theta_0$.

Subject to some regularity conditions, a locally optimum test for (24) is given by

$$\varphi_{lo}(\underline{x}) = \begin{cases} 1 \; , & \mathfrak{I}(\underline{x}) > \tau \\ \gamma \; , & \mathfrak{I}(\underline{x}) = \tau \\ 0 \; , & \mathfrak{I}(\underline{x}) < \tau \end{cases} \tag{25}$$

where

$$\mathfrak{I}(\underline{x}) = \partial \mathcal{L}_\theta(\underline{x}) / \partial\theta \,|_{\theta=\theta_0} . \tag{26}$$

$\mathcal{L}_\theta$ *is the likelihood ratio between* H_θ and H_{θ_0}. It follows from the independence assumption that

$$\mathcal{L}_\theta(\underline{x}) = \prod_{i=1}^{n} L_\theta^{(i)}(x_i)$$

where (27)

$$L_\theta^{(i)} = dP_\theta^{(i)} / \, dP_{\theta_0}^{(i)} \; ; \; i=1,\ldots,n .$$

Therefore, the locally-optimum test statistic is given by

$$\mathfrak{I}(\underline{x}) = \sum_{i=1}^{n} T^{(i)}(x_i)$$

where (28)

$$T^{(i)} = \partial L_\theta^{(i)} / \partial\theta \,|_{\theta=\theta_0} \; ; \; i=1,\ldots,n .$$

The objective is to choose a sequence $\{Q^{(i)}\}_{i=1}^{n} \equiv \underline{Q}$ of M-level quantizers to replace the nonlinearities $T^{(i)}$ of Equation (28). Then, the new test will be the following

$$\varphi_{\underline{Q}}(\underline{x}) = \begin{cases} 1, & \Im_{\underline{Q}}(\underline{x}) > \tau' \\ \gamma', & \Im_{\underline{Q}}(\underline{x}) = \tau' \\ 0, & \Im_{\underline{Q}}(\underline{x}) < \tau' \end{cases} \tag{29}$$

where

$$\Im_{\underline{Q}}(\underline{x}) = \sum_{i=1}^{n} Q^{(i)}(x_i).$$

For each i, $Q^{(i)}$ has M levels and can be represented as a pair $(\underline{t}^{(i)}, \underline{q}^{(i)})$, where $\underline{q}^{(i)} \in \mathbb{R}^M$ are the levels of $Q^{(i)}$. The breakpoints $\underline{t}^{(i)}$ are such that $-\infty = t_0^{(i)} < t_1^{(i)} < \ldots < t_{M-1}^{(i)} < t_M(i) = \infty$, and we take $Q^{(i)}(x) = q_k^{(i)}$ when $x \in (t_{k-1}^{(i)}, t_k^{(i)}]$, $k=1,\ldots,M$.

Note that the notation here is different than Kassam's.

I. Fixed Breakpoints - Locally Optimum Quantization

The first step taken by Poor and Thomas is the same taken by Kassam. That is the optimum sequence of level vectors $\{\underline{q}^{(i)}\}_{i=1}^{n}$ was derived for a fixed sequence of breakpoint vectors $\{\underline{t}^{(i)}\}_{i=1}^{n}$. For fixed $\{\underline{t}^{(i)}\}_{i=1}^{n}$, the post-quantization likelihood ratio is given by

$$\mathcal{L}_{\theta}^{Q}(\underline{x}) = \prod_{i=1}^{n} \{P_{\theta}^{(i)}(t_{k_i-1}^{(i)}, t_{k_i}^{(i)}]/P_{\theta_0}^{(i)}(t_{k_i-1}^{(i)}, t_{k_i}^{(i)}]\} \tag{30}$$

$$\text{for} \quad x_i \in (t_{k_i-1}^{(i)}, t_{k_i}^{(i)}] \; ; \; i=1,\ldots,n.$$

The locally-optimum (post-quantization) statistic is given by

$$\partial \mathcal{L}_{\theta}^{Q}(\underline{x})/\partial\theta\big|_{\theta=\theta_0} = \sum_{i=1}^{n} \{(\partial P_{\theta}^{(i)}(t_{k_i-1}^{(i)}, t_{k_i}^{(i)}]/\partial\theta\big|_{\theta=\theta_0})/P_{\theta_0}(t_{k_i-1}^{(i)}, t_{k_i}^{(i)}]\}. \tag{31}$$

On the other hand, the test statistic to be used is of the form

$$\mathfrak{I}_{\underline{Q}}(\underline{x}) = \sum_{i=1}^{n} Q^{(i)}(x_i). \tag{32}$$

The authors note that

$$\partial \mathcal{L}_{\theta}^{\underline{Q}}(\underline{x})/\partial\theta\big|_{\theta=\theta_0} = \mathfrak{I}_{\underline{Q}}(\underline{x}).$$

if the level vectors are chosen to be

$$q_k^{(i)} = [\partial P_{\theta}^{(i)}(t_{k-1}^{(i)}, t_k^{(i)}]/\partial\theta\big|_{\theta=\theta_0}]/P_{\theta_0}^{(i)}(t_{k-1}^{(i)}, t_k^{(i)}) \tag{33}$$

for k=1,...,M and i=1,...,n .

Thus, the choice of Equation (33) allows the post-quantization test to be represented optimally in the form of Equation (29). Under a regularity condition, (33) becomes

$$q_k^{(i)} = \int_{t_{k-1}^{(i)}}^{t_k^{(i)}} T^{(i)} dP_{\theta_0}^{(i)} \Big/ \int_{t_{k-1}^{(i)}}^{t_k^{(i)}} dP_{\theta_0}^{(i)} \quad ; \; k=1,\ldots,M \text{ and } i=1,\ldots,n. \tag{34}$$

Therefore, the problem of locally-optimum quantization is again reduced to that of optimally selecting the breakpoint vectors $\{\underline{t}^{(i)}\}_{i=1}^{n}$. Note that Equation (34) is a necessary and sufficient condition, because its validity is based only on the generalized Neyman-Pearson lemma and not on the existence of a stationary point. (The latter was true in Kassam's approach).

II. Asymptotically Optimum Choice of Breakpoints

For the asymptotic case ($n \to \infty$), and under some further regularity conditions, Poor and Thomas show that for the sequence $\underline{Q}$ to be optimum, $\underline{t}^{(i)}$ must be chosen to maximize $\mathrm{Var}_{\theta_0}(Q^{(i)})$ for each i,

where $$\mathrm{Var}_{\theta_0}(Q^{(i)}) = \sum_{k=1}^{M} \int_{t_{k-1}^{(i)}}^{t_k^{(i)}} T^{(i)} dP_{\theta_0}^{(i)} \Big/ \int_{t_{k-1}^{(i)}}^{t_k^{(i)}} dP_{\theta_0}^{(i)}. \tag{35}$$

To search for the maxima of Equation (35), $\underset{\underline{t}^{(i)}}{\mathrm{grad}}\ \mathrm{Var}_{\theta_0}(Q^{(i)})$ is set equal to zero. This yields necessary conditions to be satisfied by the optimum $\underline{t}^{(i)}$; namely

$$\text{(a)}\quad T^{(i)}(t_k^{(i)}) = (q_k^{(i)} + q_{k+1}^{(i)})/2 \ ;\ k=1,\ldots,(M-1)$$

where (from eq. (34)) (36)

$$\text{(b)}\quad q_k^{(i)} = \int_{t_{k-1}^{(i)}}^{t_k^{(i)}} T^{(i)} dP_{\theta_0}^{(i)} \Big/ \int_{t_{k-1}^{(i)}}^{t_k^{(i)}} dP_{\theta_0}^{(i)} \ ;\ k=1,\ldots,M.$$

The condition 36(b) is sufficient as explained above, but the same cannot be claimed about condition 36(a). The sufficiency of the latter must be checked by examining the definiteness of the matrix of second partial derivatives (Hessian matrix).

III. The Maximum-Efficacy Quantizer

The efficacy of the test $\varphi_{\underline{Q}}$ based on the sequence $\underline{Q}$ of quantizers is calculated to be (see [21])

$$\eta_{\underline{Q}} = \lim_{n\to\infty} \eta_n(\underline{Q}) \tag{37}$$

where

$$\eta_n(\underline{Q}) = \Bigg[\sum_{i=1}^{n} \sum_{k=1}^{M} q_k^{(i)} \int_{t_{k-1}^{(i)}}^{t_k^{(i)}} T^{(i)} dP_{\theta_0}^{(i)} \Bigg]^2 \Bigg/ n \Bigg[\sum_{i=1}^{n} \sum_{k=1}^{M} (q_k^{(i)})^2 \int_{t_{k-1}^{(i)}}^{t_k^{(i)}} dP_{\theta_0}^{(i)} -$$

$$- \Big(\sum_{i=1}^{n} \sum_{k=1}^{M} q_k^{(i)} \int_{t_{k-1}^{(i)}}^{t_k^{(i)}} dP_{\theta_0}^{(i)} \Big)^2 \Bigg]. \tag{38}$$

Equation (38) can be rewritten as

$$\eta_n(Q) = \mathrm{Cov}^2_{\theta_0}(\mathfrak{I}_Q,\ m(\{\underline{t}^{(i)}\}_{i=1}^{n}))/n\ \mathrm{Var}_{\theta_0}(\mathfrak{I}_Q) \tag{39}$$

where

$$m(\underline{x}\ ;\ \{\underline{t}^{(i)}\}_{i=1}^{n}) = \sum_{i=1}^{n} M^{(i)}(x_i) \tag{40}$$

with

$$M^{(i)}(x) = \int_{t_{k-1}^{(i)}}^{t_k^{(i)}} T^{(i)} dP^{(i)}_{\theta_0} \Big/ \int_{t_{k-1}^{(i)}}^{t_k^{(i)}} dP^{(i)}_{\theta_0} \tag{41}$$

for

$$x \in (t_{k-1}^{(i)},\ t_k^{(i)}],\ i=1,\ldots,M.$$

From Equation (39), by employing the Schwartz inequality, the authors *proceed to show that*

$$\eta_n(Q) \le \mathrm{Var}_{\theta_0}(m(\{\underline{t}^{(i)}\}_{i=1}^{n}))/n \tag{42}$$

with equality if and only if

$$\mathfrak{I}_Q(\underline{x}) = am(\underline{x}\ ;\ \{\underline{t}^{(i)}\}_{i=1}^{n}) + b \quad \forall x \in \mathbb{R}^n \text{ and}$$

for some numbers $a \neq 0$ and b.

Thus, if the choice

$$\mathfrak{I}_Q(\underline{x}) = m(\underline{x}\ ;\ \{\underline{t}^{(i)}\}_{i=1}^{n}) \tag{43}$$

is made, for $\{\underline{t}^{(i)}\}_{i=1}^{n}$ given, the maximum efficacy will be achieved.

Equations (40) and (43) imply that, for fixed $\underline{t}^{(i)}$, the optimum choice of level vectors is given by

$$q_k^{(i)} = \int_{t_{k-1}^{(i)}}^{t_k^{(i)}} T^{(i)} dP^{(i)}_{\theta_0} \Big/ \int_{t_{k-1}^{(i)}}^{t_k^{(i)}} dP^{(i)}_{\theta_0};\ k=1,\ldots,M;\ i=1,\ldots,\infty. \tag{44}$$

Equation (44) is identical to Equation (34). Therefore, the authors conclude, for fixed $\{\underline{t}^{(i)}\}_{i=1}^{\infty}$, this choice of levels is both locally-optimum and asymptotically-most-efficient.

Applying Eq. (44) to the expression for $\eta_n(Q)$ (Eq. 38) one has

$$\eta_n(Q) = \sum_{i=1}^{n} \mathrm{Var}_{\theta_0}(Q^{(i)})/n \ . \tag{45}$$

Hence, the maximum-efficacy quantizer $Q^{(i)}$ maximizes $\mathrm{Var}_{\theta_0}(Q^{(i)})$ and the set of Equations (10)(a),(b) gives the necessary conditions for this case as well.

Again, note that Equation (44) is sufficient since it follows from the Schwartz inequality. Recall that in Kassam's problem, he based his results on the existence of stationary points and the sufficiency of his corresponding equation is not guaranteed.

IV. The Optimum Quantizer as an Approximation to the Locally Optimum Nonlinearity - General Case

In their treatment of the general case, Poor and Thomas did not consider the MMSE between quantized data and data transformed by the locally-optimum nonlinearity as a criterion for optimum quantization. We have seen previously that Kassam, examining the above criterion, showed that it leads to the same set of necessary conditions for the parameters of the optimum quantizer as the criterion of maximum efficacy. We will now extend Kassam's findings to the general case.

Adhering to the same notation as in the previous section, the optimum test statistic, in the general case, is given by

$$\mathfrak{J}(\underline{x}) = \sum_{i=1}^{n} T^{(i)}(x_i).$$

The post-quantization test statistic is given by

$$\mathfrak{I}_{Q}(\underline{x}) = \sum_{i=1}^{n} Q^{(i)}(x_i) .$$

Let $\epsilon^{(i)}$ denote the mean-squared-error between the two quantities of interest

$$\epsilon^{(i)} = E_{\theta_0}\{[Q^{(i)} - T^{(i)}]^2\} \tag{46}$$

$$\Rightarrow \epsilon^{(i)} = E_{\theta_0}\{(Q^{(i)})^2\} + E_{\theta_0}\{(T^{(i)})^2\} - 2E_{\theta_0}\{T^{(i)}Q^{(i)}\}$$

$$\Rightarrow \epsilon^{(i)} = E_{\theta_0}\{(Q^{(i)})^2\} - 2E_{\theta_0}\{T^{(i)}Q^{(i)}\} + I_{\theta_0}^{(i)} \tag{47}$$

where $$I_{\theta_0}^{(i)} = E_{\theta_0}\{(T^{(i)})^2\} = E_{\theta_0}\{[\partial L_{\theta}^{(i)}/\partial\theta|_{\theta=\theta_0}]^2\} .$$

Obviously, $\epsilon^{(i)}$ will not be finite unless $I_{\theta_0}^{(i)}$ is finite and we assume *the latter to be the case.*

$$\epsilon^{(i)} = \sum_{k=1}^{M} q_k^2(i)\int_{t_{k-1}^{(i)}}^{t_k^{(i)}} dP_{\theta_0}^{(i)} - 2\sum_{k=1}^{M} q_k^{(i)}\int_{t_{k-1}^{(i)}}^{t_k^{(i)}} T^{(i)} dP_{\theta_0}^{(i)} + I_{\theta_0}^{(i)} \quad ; \; i=1,\ldots,n . \tag{48}$$

Assuming, for the moment, that the vectors $\{\underline{t}^{(i)}\}_{i=1}^{n}$ are fixed we have

$$\partial\epsilon^{(i)}/\partial q_k^{(i)} = 0 \Rightarrow$$

$$\Rightarrow 2\sum_{k=1}^{M}\left[q_k^{(i)}\int_{t_{k-1}^{(i)}}^{t_k^{(i)}} dP_{\theta_0}^{(i)} - \int_{t_{k-1}^{(i)}}^{t_k^{(i)}} T^{(i)} dP_{\theta_0}^{(i)}\right] = 0$$

$$\Rightarrow q_k^{(i)} = \int_{t_{k-1}^{(i)}}^{t_k^{(i)}} T^{(i)} dP_{\theta_0}^{(i)} \Big/ \int_{t_{k-1}^{(i)}}^{t_k^{(i)}} dP_{\theta_0}^{(i)} \tag{49}$$

for $k=1,\ldots,M$; $i=1,\ldots,n$.

This is the same condition (eq.(44)) that the maximum-efficacy quantizer must satisfy. Kassam showed this for the additive noise, constant signal

case but he did not prove this condition to be sufficient. We will now show that equation (49) is actually a sufficient condition for the general case.

Let $\hat{Q}^{(i)}$ be a quantizer whose level vector $\underline{q}^{(i)}$ satisfies eq. (49) for a given breakpoint vector $\underline{t}^{(i)}$ and let $Q^{(i)}$ be any other quantizer with the same breakpoint vector. Consider the expression

$$\epsilon^{(i)} = E_{\theta_0}\{[Q^{(i)}-T^{(i)}]^2\}$$

$$= E_{\theta_0}\{[Q^{(i)}-\hat{Q}^{(i)}+\hat{Q}^{(i)}-T^{(i)}]^2\} =$$

$$= E_{\theta_0}\{[Q^{(i)}-\hat{Q}^{(i)}]^2\}+E_{\theta_0}\{[\hat{Q}^{(i)}-T^{(i)}]^2\}-2E_{\theta_0}\{[Q^{(i)}-\hat{Q}^{(i)}][\hat{Q}^{(i)}-T^{(i)}]\}.$$

Since $E_{\theta_0}\{[Q^{(i)}-\hat{Q}^{(i)}]^2\} \geq 0$, it would suffice to show

$$E_{\theta_0}\{[Q^{(i)}-\hat{Q}^{(i)}][\hat{Q}^{(i)}-T^{(i)}] = 0 \quad ,$$

in order to prove that $Q^{(i)}$ is the quantizer yielding the lowest MMSE between the quantities considered, from among all M-level quantizers with the same fixed breakpoints. We have

$$E_{\theta_0}\{[Q^{(i)}-\hat{Q}^{(i)}][\hat{Q}^{(i)}-T^{(i)}]\} =$$

$$= E_{\theta_0}\{Q^{(i)}\hat{Q}^{(i)}\}-E_{\theta_0}\{\hat{Q}^{(i)}T^{(i)}\}-E_{\theta_0}\{Q^{(i)}T^{(i)}\}+E_{\theta_0}\{[\hat{Q}^{(i)}]^2\}$$

$$= \sum_{k=1}^{M}\left[q_k^{(i)}\hat{q}_k^{(i)}\int_{t_{k-1}^{(i)}}^{t_k^{(i)}} dP_{\theta_0}^{(i)} - \hat{q}_k^{(i)}\int_{t_{k-1}^{(i)}}^{t_k^{(i)}} T^{(i)} dP_{\theta_0}^{(i)} - q_k^{(i)}\int_{t_{k-1}^{(i)}}^{t_k^{(i)}} T^{(i)} dP_{\theta_0} + (\hat{q}_k^{(i)})^2\int_{t_{k-1}^{(i)}}^{t_k^{(i)}} dP_{\theta_0}^{(i)}\right].$$

Substituting for $\hat{q}_k^{(i)}$ from equation (49) we have

$$E_{\theta_0}\{[Q^{(i)}-\hat{Q}^{(i)}][\hat{Q}^{(i)}-T^{(i)}]\} =$$

$$= \sum_{k=1}^{M} \left\{ q_k^{(i)} \int_{t_{k-1}^{(i)}}^{t_k^{(i)}} T^{(i)} dP_{\theta_0}^{(i)} - \left[\int_{t_{k-1}^{(i)}}^{t_k^{(i)}} T^{(i)} dP_{\theta_0}^{(i)} \right]^2 \Big/ \int_{t_{k-1}^{(i)}}^{t_k^{(i)}} dP_{\theta_0}^{(i)} - \right.$$

$$\left. - q_k^{(i)} \int_{t_{k-1}^{(i)}}^{t_k^{(i)}} T^{(i)} dP_{\theta_0}^{(i)} + \left[\int_{t_{k-1}^{(i)}}^{t_k^{(i)}} T^{(i)} dP_{\theta_0}^{(i)} \right]^2 \Big/ \int_{t_{k-1}^{(i)}}^{t_k^{(i)}} dP_{\theta_0}^{(i)} \right\} =$$

$$= 0.$$

Hence, we conclude that $\hat{Q}^{(i)}$ yields the lowest value for $\epsilon^{(i)}$ from among all M-level quantizers for a fixed $\underline{t}^{(i)}$ and, in essence, that Equation (49) is sufficient for minimum $\epsilon^{(i)}$.

Now, substituting for $q_k^{(i)}$ as defined by Eq. (49) into the expression for $\epsilon^{(i)}$ given by Eq. (48) we have

$$\epsilon^{(i)} = \sum_{k=1}^{M} \left\{ \left[\int_{t_{k-1}^{(i)}}^{t_k^{(i)}} T^{(i)} dP_{\theta_0}^{(i)} \right]^2 \Big/ \int_{t_{k-1}^{(i)}}^{t_k^{(i)}} dP_{\theta_0}^{(i)} - 2 \left[\int_{t_{k-1}^{(i)}}^{t_k^{(i)}} T^{(i)} dP_{\theta_0}^{(i)} \right]^2 \Big/ \int_{t_{k-1}^{(i)}}^{t_k^{(i)}} dP_{\theta_0}^{(i)} \right\} + I_{\theta_0}^{(i)}$$

$$= I_{\theta_0}^{(i)} - \sum_{k=1}^{M} \left[\int_{t_{k-1}^{(i)}}^{t_k^{(i)}} T^{(i)} dP_{\theta_0}^{(i)} \right]^2 \Big/ \int_{t_{k-1}^{(i)}}^{t_k^{(i)}} dP_{\theta_0}^{(i)}$$

$$= I_{\theta_0}^{(i)} - \mathrm{Var}_{\theta_0}(Q^{(i)}) \; ; \; i=1,\ldots,n. \qquad (50)$$

It is clear now that minimizing $\epsilon^{(i)}$ is equivalent to maximizing $\mathrm{Var}_{\theta_0}(Q^{(i)})$, which is exactly the same condition that must be satisfied by the maximum-efficacy quantizer, when the choice of Eq. (49) for the elements of the level vectors $\underline{q}^{(i)}$ is made. Since Equation (49) was proven to be sufficient and Equation (44) was shown by Poor and Thomas also to be sufficient, it is safe to claim that the criterion of MMSE between quantized data and data transformed by the locally optimum non-linearity and the one of maximum efficacy are completely equivalent.

That is, every time $\epsilon^{(i)}$ is at a minimum for each i, $\eta_{\underline{Q}}$ achieves a maximum.

V. A Condition for Sufficiency

We have noted that the sufficiency of Equation (36) (a) remained to be tested. The straightforward way to test its sufficiency would be by examining the definiteness of the matrix of the second partial derivatives of either the efficacy or the MSE expression. We have attempted here to carry out this procedure for the additive noise, constant signal case (Kassam's case). The resulting expressions for the matrix elements, however, were not strictly positive or negative and obviously, not every noise distribution will lead to the sufficiency of the equation in question. We were unable to find a condition on the noise density which would guarantee its sufficiency. The expressions for the matrix elements can be found in the Appendix.

Another approach, based on Fleischer's [22] work, led us to a condition on the noise statistics which, when satisfied, leads to sufficiency.

(a). Sufficient Conditions for Minimum Distortion

In the opening section we stated that a criterion of quantizer performance widely accepted is the one of minimum distortion D, where

$$D = \sum_{k=1}^{M} \int_{t_{k-1}}^{t_k} (x-q_k)^2 f(x)dx \,. \tag{51}$$

We have also pointed-out that Max [1] derived the following necessary conditions that the parameters of the minimum-distortion quantizer must satisfy.

$$q_k = \frac{\int_{t_{k-1}}^{t_k} xf(x)dx}{\int_{t_{k-1}}^{t_k} f(x)dx} \quad ; \; k=1,\ldots,M \tag{52}$$

$$t_k = \tfrac{1}{2}(q_k + q_{k+1}) \quad ; \; k=1,\ldots,(M-1) \; . \tag{53}$$

By examining the matrix of the second partials, Fleischer showed that this set of conditions is sufficient, if the noise p.d.f. f(x) obeys the relation

$$\frac{d^2}{dx^2}\,[\ln f(x)] < 0 \tag{54}$$

This is equivalent to

$$-\frac{d}{dx}[\ln f(x)] = -\frac{f'(x)}{f(x)} = \text{increasing}$$

or

$$-\ln f(x) = \text{convex} \; .$$

A density satisfying eq. (54) is known as <u>strongly unimodal</u>. We will apply this result to derive a condition for the sufficiency of the remaining necessary condition (eq. (36),(a)) for the optimum detection quantizer.

(b). A Sufficient Condition for Optimum Detection

Consider the expression

$$D^{(i)} = E_{\theta_0}\{[Q^{(i)}(x) - Y]^2\} \tag{55}$$

where

$$Y = T^{(i)}(x)$$

and $Q^{(i)}$ is an M-level quantizer. The notation is again the same used in the treatment of the general case. Recall that minimization of $D^{(i)}$, as defined here, results in optimum detection. We have

$$D^{(i)} = \int_{t_{k-1}^{(i)}}^{t_k^{(i)}} (Q^{(i)}(x) - Y)^2 \, dP_{\theta_0} \; ; \; k=1,\ldots,M \, .$$

Let h(Y) denote the p.d.f. of Y, assuming this p.d.f. exists. Then, a change of variables leads to

$$D^{(i)} = \int_{t'^{(i)}_{k-1}}^{t'^{(i)}_k} (Q'^{(i)}(y)-y)^2 \, h(y)dy \; ; \; k=1,\ldots,M$$

where

$$t'^{(i)}_k = T^{(i)}(t_k^{(i)}) \; ; \; k=1,\ldots,M \, .$$

Then, Max's necessary conditions for minimum $D^{(i)}$ become

$$q'^{(i)}_k = \frac{\int_{t'^{(i)}_{k-1}}^{t'^{(i)}_k} yh(y)dy}{\int_{t'^{(i)}_{k-1}}^{t'^{(i)}_k} h(y)dy} \; ; \; k=1,\ldots,M \tag{56}$$

$$t'^{(i)}_k = \tfrac{1}{2}(q'^{(i)}_k + q'^{(i)}_{k+1}) \; ; \; k=1,\ldots,(M-1). \tag{57}$$

and Fleischer's condition for sufficiency of the above equations is

$$-\ln h(y) = \text{convex}. \tag{58}$$

If $T^{(i)}(\cdot)$ is invertible, then the parameters of the optimum quantizer $Q^{(i)}(x)$ can be defined by

$$\left.\begin{aligned} t_k^{(i)} &= T^{-1(i)}(t'^{(i)}_k) \\ q_k^{(i)} &= T^{-1(i)}(q'^{(i)}_k). \end{aligned}\right\} \; , \; k=1,\ldots,M$$

In general, if $x_1,\ldots,x_n$ are all the real roots of the equation

$$y = T^{(i)}(x)$$

then the p.d.f. h(y) is given by

$$h(y) = \frac{f(x_1)}{|T'^{(i)}(x_1)|} + \cdots\cdots + \frac{f(x_n)}{|T'^{(i)}(x_n)|}$$

where

$$T'^{(i)}(x) = dT^{(i)}(x)/dx.$$

For the case at hand ($T^{(i)}(x)$ invertible), only one real root exists and we have

$$h(y) = \frac{f(x)}{|T'^{(i)}(x)|}. \tag{59}$$

3. APPLICATIONS TO SIGNAL DETECTION

In this section we will examine the optimum-quantizer structure for several cases of practical interest resulting from the general case of local decisions. These examples can be found in [21].

A. Known Signals in Additive Noise (Kassam's Case)

For this case, the hypothesis pair reduces to

$$H_{\theta_0} : x_i \sim f(x) \; ; \; i=1,\ldots,n$$

vs

$$H_{\theta} : x_i \sim f(x-\theta S_i) \; ; \; i=1,\ldots,n .$$

$\{S_i\}_{i=1}^{n}$ is a known signal sequence and θ is assumed to be positive. Again, for the local case, consider $\theta \to 0^+$.

The i^{th} likelihood ratio is

$$L_{\theta}^{(i)}(x) = f(x-\theta S_i)/f(x).$$

By differentiating, we obtain

$$T^{(i)}(x) = -S_i f'(x)/f(x) .$$

From Eq. (36), the optimum quantizer sequence is given by

$$Q^{(i)}(x) = S_i Q(x)$$

where $Q=(\underline{t},\underline{q})$ is the solution to

$$\begin{aligned} &(a) \quad -f'(t_k)/f(t_k) = (q_k+q_{k+1})/2 \; ; \; k=1,\ldots,(M-1) \\ &(b) \quad q_k = [f(t_{k-1})-f(t_k)]/\int_{t_{k-1}}^{t_k} f(x)dx \; ; \; k=1,\ldots,M . \end{aligned} \tag{60}$$

The above, is exactly the same set of conditions derived by Kassam. One important difference is the fact that Eq. (16)(b) is now a sufficient condition and only the sufficiency of Eq. (16)(a) remains to be tested.

B. Stochastic Signals in Noise (Additive-Noise Model)

For this case, the hypotheses pair is given by

$$H_{\theta_0} : x_i \sim f(x) \; ; \; i=1,\ldots,n$$

vs

$$H_\theta : x_i \sim \int_{-\infty}^{\infty} f(x-\theta^{\frac{1}{2}}s)dG_i(s) \; ; \; i=1,\ldots,n.$$

$\{G_i\}_{i=1}^n$ is a sequence of zero-mean distribution functions corresponding to a sequence of independent samples from a stochastic signal. Again assume $\theta>0$ and $\theta \to 0^+$. The i^{th} likelihood ratio is

$$L_\theta^{(i)}(x) = \int_{-\infty}^{\infty} f(x-\theta^{\frac{1}{2}}s)dG_i(s)/f(x)$$

and i^{th} locally-optimum nonlinearity is

$$T^{(i)}(x) = \sigma_i^2 f''(x)/2f(x)$$

where

$$\sigma_i^2 = \int_{-\infty}^{\infty} s^2 dG_i(s).$$

For this case, the optimum quantizer sequence is given by

$$Q^{(i)}(x) = \sigma_i^2 Q(x)/2$$

where $Q \equiv (\underline{t},\underline{q})$ is the solution to

$$\begin{aligned} &\text{(a)} \quad f''(t_k)/f(t_k) = (q_k + q_{k+1})/2 \; ; \; k=1,\ldots,(M-1) \\ &\text{(b)} \quad q_k = [f'(t_k)-f'(t_{k-1})]/\int_{t_{k-1}}^{t_k} f(x)dx \; ; \; k=1,\ldots,M. \end{aligned} \tag{61}$$

C. Stochastic Signals in Noise (Scale-Change Model)

For this case, we have

$$H_{\theta_0} : X_i \sim f(x) \; ; \; i=1,\ldots,n$$

vs

$$H_\theta : X_i \sim f(x/v_i)/v_i \; ; \; i=1,\ldots,n$$

where

$$v_i = [1 + \theta\sigma_i^2/\sigma^2]^{\frac{1}{2}} ;$$

f is a differentiable p.d.f. with variance σ^2 and $\{\sigma_i^2\}_{i=1}^n$ a sequence of signal variances. Once more, $\theta>0$ and $\theta \to 0$. We have

$$L_\theta^{(i)}(x) = f(x/v_i)/v_i f(x)$$

and

$$T_\theta^{(i)}(x) = (\sigma_i^2/2\sigma^2)(-xf'(x)/f(x) - 1).$$

The optimum quantizer sequence is

$$Q^{(i)}(x) = (\sigma_i^2/2\sigma^2)Q(x)$$

where $Q=(\underline{t},\underline{q})$ is the solution to

$$\begin{aligned} &\text{(a)} \quad (-t_k f'(t_k)/f(t_k) - 1) = (q_k + q_{k+1})/2 \ ; \ k=1,\ldots,(M-1) \\ &\text{(b)} \quad q_k = [t_{k-1}f(t_{k-1}) - t_k f(t_k)]/\int_{t_{k-1}}^{t_k} f(x)dx . \end{aligned} \tag{62}$$

4. NUMERICAL RESULTS

The set of simultaneous equations for the optimum parameters can be solved by several different numerical methods. One such iterative numerical technique is described by Max [1]. Using a technique similar to that, Kassam [20] obtained the optimum quantizer parameters for the additive noise, known signal case and for noise densities in the class of generalized Gaussian noise densities which contains a wide range of non-Gaussian p.d.f's, parametrized by their rates of exponential decay. A generalized Gaussian density $f_p(x)$ is defined by

$$f_p(x) = \frac{p}{2\Gamma(1/p)A(p)} \exp\{-[|x|/A(p)]^p\} \quad , p>0 \tag{63}$$

where $$A(p) = [\sigma^2\Gamma(1/p)/\Gamma(3/p)]^{\frac{1}{2}}.$$

$\Gamma(\cdot)$ is the gamma function and σ^2 the variance of the density. Note that p=2 produces the Gaussian density with variance σ^2. For the rest of this discussion unit variance is assumed, that is $\sigma^2=1$.

Kassam evaluated the optimum quantizer parameters under the locally optimum detection criterion as well as under the minimum squared-error distortion criterion (Max's quantizer) for densities of the general class described above. This procedure was carried out for the cases of four-level and eight-level quantization (m=2 and m=4 respectively in Kassam's (2m)-level symmetric quantizer). The results show, as it might be expected, that the locally-optimum quantizer produces high efficacy and high distortion while the minimum-distortion quantizer results in both low distortion and low detection efficacy.

By using a different numerical method we have exactly duplicated Kassam's results for m=4. This was done basically in order to test our program. We will now produce analogous results for the stochastic signal

case (both additive-noise and scale-change models)

A. General Procedure

It was shown by Kassam that for symmetric noise densities the optimum quantizer (for his case) is odd symmetric. For stochastic signals, the optimum nonlinearities are even symmetric and the corresponding optimum quantizers will also be even symmetric. Since the generalized Gaussian is a symmetric density, we will be looking for (even) symmetric quantizers. The notation to be used in this section is the following.

The positive input values are partitioned into m intervals $T_1,\ldots,T_m$ where $T_k=(t_{k-1},t_k]$ and $\{t_k\}_{k=0}^{m}$ is an increasing sequence of non-negative numbers with $t_0=0$ and $t_m=\infty$. The output level corresponding to T_k is denoted by q_k. The definitions $T_{-k}=[-t_k,-t_{k-1})$ and $q_{-k}=q_k$ for $1\le k\le m$ complete the specification of the symmetric quantizer.

The basis of our procedure is a program that uses the Davidon-Fletcher-Powell [23] (DFP) algorithm to minimize a function of n variables. The program produces the minimum value of the function as well as the values of the variables that lead to the minimum value. An individual subroutine is needed to provide the value of the function at hand as well as the gradient vector for each input vector.

B. The Minimum Distortion Quantizer

Max's quantizer is designed to minimize the distortion D which, in terms of the quantization parameters, is

$$D = \sum_{k=1}^{M}\left[q_k^2\int_{t_{k-1}}^{t_k} f(x)dx + \int_{t_{k-1}}^{t_k} x^2 f(x)dx - 2q_k\int_{t_{k-1}}^{t_k} f(x)dx\right]. \qquad (64)$$

The optimum level values are given by

$$q_k = \frac{\int_{t_{k-1}}^{t_k} xf(x)dx}{\int_{t_{k-1}}^{t_k} f(x)dx} \tag{65}$$

and the gradient vector is defined by

$$g_k = 2t_k f(t_k)(q_{k+1}-q_k) + f(t_k)(q_k^2-q_{k+1}^2) \;;\; k=1,\ldots,(M-1). \tag{66}$$

The DFP program is used to minimize D, after the optimum values for q_k are substituted in (from eq. (65)). Then the algorithm produces the optimum values for the breakpoint vector $\underline{t}$.

C. Stochastic Signal - Additive Noise

After substituting for q_k in the efficacy expression (Eq. (38)) by means of Eq. (61)(b), the following expression must be maximized

$$\mathrm{Var}_0(Q) = \sum_{k=1}^{M} [f'(t_k)-f'(t_{k-1})]^2 / \int_{t_{k-1}}^{t_k} f(x)dx. \tag{67}$$

Then the minimizing program is used to minimize $-\mathrm{Var}_0(Q)$ and to produce the optimum breakpoints. Eq. (61)(b) will then produce the optimum levels.

In order to determine the efficacy value produced by Max's quantizer, the parameters $(\underline{t},\underline{q})$ determined as described in the previous section, must be substituted directly into the general efficacy expression given by (see Eq. (38))

$$\eta(Q) = \left[\sum_{k=1}^{M} q_k [f'(t_k)-f'(t_{k-1})]\right]^2 / \left[\sum_{k=1}^{M} q_k^2 \int_{t_{k-1}}^{t_k} f(x)dx - \left(\sum_{k=1}^{M} q_k \int_{t_{k-1}}^{t_k} f(x)dx\right)^2\right]. \tag{68}$$

D. Stochastic Signal - Scale Change

The same basic procedure is followed in this case as in the previous section. Here, the efficacy expression is given by

$$\eta(Q)=\left[\sum_{k=1}^{M} q_k\left[t_{k-1}f(t_{k-1})-t_k f(t_k)\right]\right]^2 / \left[\sum_{k=1}^{M} q_k^2 \int_{t_{k-1}}^{t_k} f(x)dx - \left(\sum_{k=1}^{M} q_k \int_{t_{k-1}}^{t_k} f(x)dx\right)^2\right] \quad (69)$$

and substituting for q_k from Eq. (62)(b), Eq. (69) becomes

$$\eta(Q)=\mathrm{Var}_0(Q)=\sum_{k=1}^{M}\left[t_{k-1}f(t_{k-1})-t_k f(t_k)\right]^2 / \int_{t_{k-1}}^{t_k} f(x)dx\,. \quad (70)$$

E. Tables - Graphs - Discussion

TABLE I

Parameters of MMSE quantizer (Max's quantizer) generalized Gaussian density, m=4.

p	1.0	1.2	1.4	1.6	1.8	2.0	2.2	2.4	2.6	2.8	3.0
q_1	0.233	0.239	0.243	0.244	0.245	0.245	0.245	0.244	0.243	0.242	0.241
q_2	0.833	0.807	0.789	0.775	0.765	0.756	0.749	0.742	0.737	0.732	0.728
q_3	1.673	1.557	1.478	1.422	1.378	1.344	1.317	1.294	1.275	1.260	1.246
q_4	3.087	2.751	2.526	2.366	2.246	2.152	2.077	2.017	1.966	1.923	1.887
t_1	0.533	0.523	0.516	0.510	0.505	0.501	0.497	0.493	0.490	0.487	0.485
t_2	1.253	1.182	1.134	1.099	1.071	1.050	1.033	1.018	1.006	0.996	0.987
t_3	2.380	2.154	2.002	1.894	1.812	1.748	1.697	1.656	1.621	1.591	1.566

TABLE II

Parameters of locally-optimum quantizer (stochastic signal-additive noise).

p	1.6	1.8	2.0	2.2	2.4	2.6	2.8	3.0
q_1	-6.369	-1.035	-0.731	-0.617	-0.560	-0.530	-0.514	-0.507
q_2	-1.262	0.087	0.557	0.950	1.355	1.797	2.289	2.835
q_3	0.288	1.653	2.562	3.527	4.631	5.895	7.345	8.983
q_4	2.319	4.265	6.136	8.325	10.930	13.976	17.518	21.568
t_1	0.014	0.581	0.956	1.194	1.362	1.483	1.574	1.643
t_2	0.430	1.268	1.600	1.790	1.911	1.991	2.044	2.080
t_3	1.355	2.070	2.313	2.428	2.486	2.511	2.519	2.515

TABLE III

Parameters of locally-optimum quantizer (stochastic signal - scale change).

p	1.0	1.2	1.4	1.6	1.8	2.0	2.2	2.4	2.6	2.8	3.0
q_1	-0.670	-0.688	-0.703	-0.714	-0.723	-0.731	-0.737	-0.743	-0.748	-0.752	-0.756
q_2	0.178	0.251	0.326	0.403	0.481	0.558	0.637	0.715	0.795	0.874	0.954
q_3	1.365	1.606	1.846	2.085	2.326	2.563	2.804	3.040	3.279	3.517	3.756
q_4	3.365	3.925	4.481	5.035	5.585	6.137	6.689	7.234	7.782	8.330	8.877
t_1	0.533	0.645	0.740	0.823	0.894	0.956	1.010	1.058	1.100	1.137	1.171
t_2	1.253	1.368	1.453	1.515	1.564	1.600	1.630	1.652	1.671	1.686	1.699
t_3	2.380	2.390	2.380	2.360	2.337	2.313	2.289	2.266	2.243	2.223	2.203

In the tables to follow, $Q_M^2(\cdot)$ will denote a quantizer with the same breakpoints as Max's (MMSE) quantizer but with level vectors equal to the square of the corresponding levels of the MMSE quantizer. The efficacy produced by $Q_M^2(\cdot)$ has also been calculated for both cases of interest. In addition, the efficacy of the locally-optimum (unquantized) detector has been calculated. The expressions for the optimum efficacies for the additive noise and scale-change models respectively and for generalized Gaussian noise densities are given by [28]

$$\mathrm{Var}_0(g_{1o}) = \frac{p^4 \eta^4(p) \Gamma(2-3/p)(1-1/p)(3-4/p)}{\Gamma(1/p)} \qquad (71)$$

$$p > 1.5$$

where $\eta(p) = [\Gamma(3/p)/\Gamma(1/p)]^{\frac{1}{2}}$

and

$$\mathrm{Var}_0(g_{1o}) = p \; ; \; p>0. \qquad (72)$$

TABLE IV

Comparison of MMSE and efficacy of locally-optimum and MMSE quantizers. Efficacies of locally-optimum detector and $Q_M^2(\cdot)$ quantizer. Stochastic signal-additive noise model.

P	LOCALLY OPTIMUM QUANTIZATION		MINIMUM DISTORTION (MMSE) QUANTIZATION		$Q_M^2(\cdot)$	LOCALLY-OPTIMUM DETECTOR
	EFFICACY	MSE	EFFICACY	MSE	EFFICACY	EFFICACY
1.6	2.018	1.524	1.767	0.040	1.505	2.851
1.8	1.680	1.178	1.645	0.037	1.605	1.932
2.0	1.788	1.370	1.526	0.035	1.667	2.000
2.2	2.044	1.680	1.414	0.033	1.698	2.282
2.4	2,408	2.076	1.311	0.031	1.705	2.698
2.6	2.874	2.558	1.215	0.030	1.691	3.234
2.8	3.446	3.028	1.125	0.029	1.659	3.893
3.0	4.128	3.680	1.044	0.028	1.616	4.681

TABLE V

Comparison of MMSE and efficacy of locally-optimum and MMSE quantizers. Efficacies of locally-optimum detector and $Q^2(\cdot)$ quantizer. Stochastic signal-scale change model.

P	LOCALLY OPTIMUM QUANTIZATION		MINIMUM DISTORTION (MMSE) QUANTIZATION		$Q_M^2(\cdot)$	LOCALLY-OPTIMUM DETECTOR
	EFFICACY	MSE	EFFICACY	MSE	EFFICACY	EFFICACY
1.0	0.892	0.632	0.891	0.054	0.783	1.000
1.2	1.070	0.754	1.058	0.048	0.996	1.200
1.4	1.250	0.896	1.172	0.043	1.191	1.400
1.6	1.430	1.048	1.328	0.040	1.368	1.600
1.8	1.608	1.206	1.434	0.037	1.526	1.800
2.0	1.788	1.370	1.526	0.035	1.667	2.000
2.2	1.968	1.538	1.605	0.033	1.793	2.200
2.4	2.148	1.726	1.673	0.031	1.904	2.400
2.6	2.328	1.802	1.730	0.030	2.002	2.600
2.8	2.508	2.054	1.779	0.029	2.088	2.800
3.0	2.688	2.232	1.822	0.028	2.165	3.000

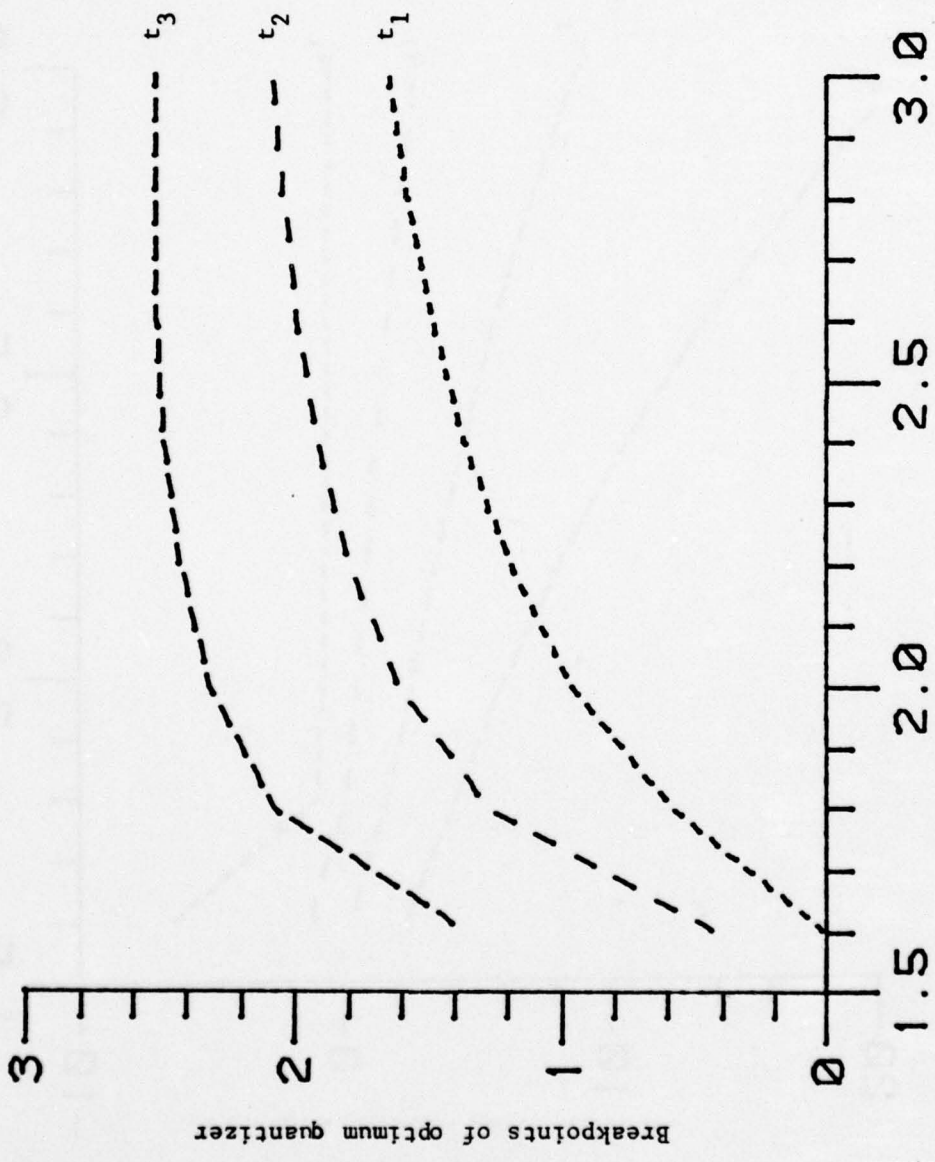


Figure 1. Breakpoints of 4-level locally-optimum quantizer. Stochastic signal, additive noise model.

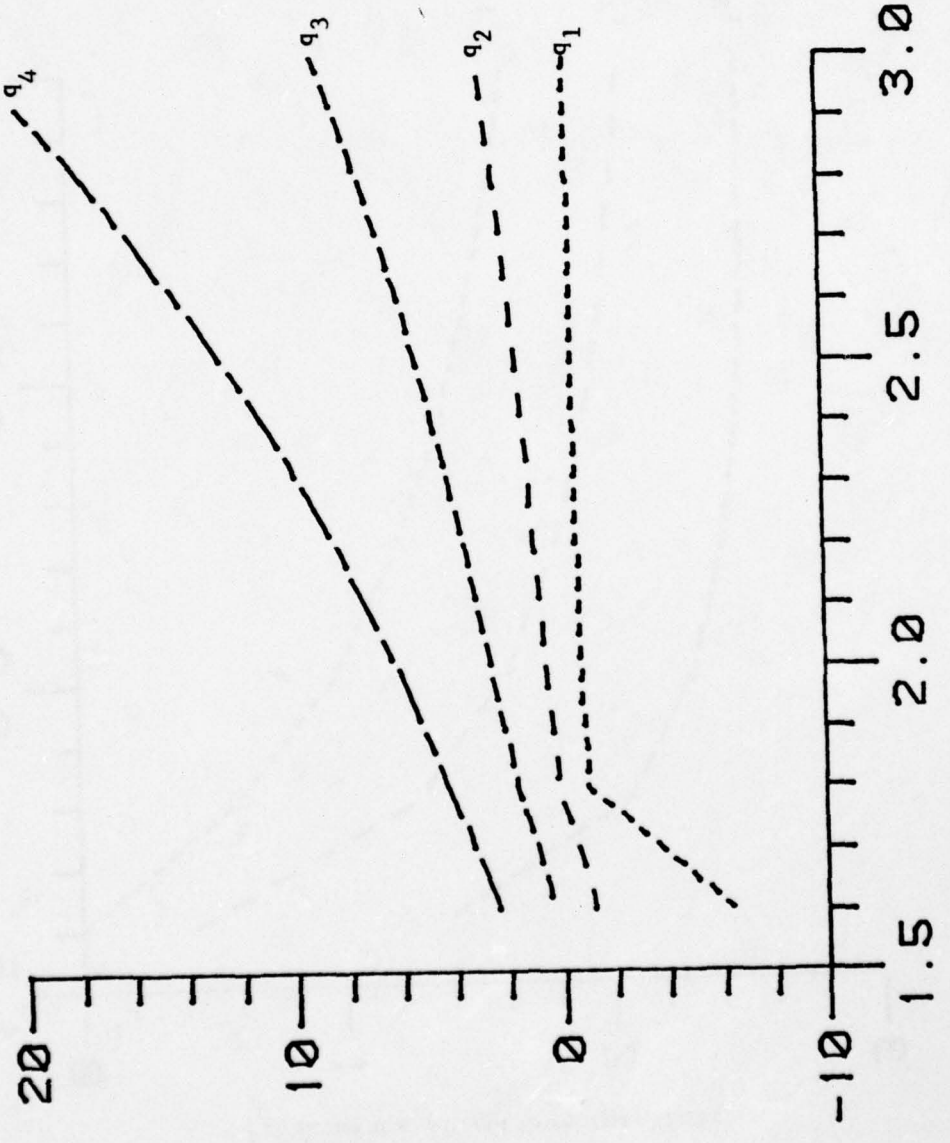


Figure 2. Levels of 4-level locally-optimum quantizer. Stochastic signal, additive noise model.

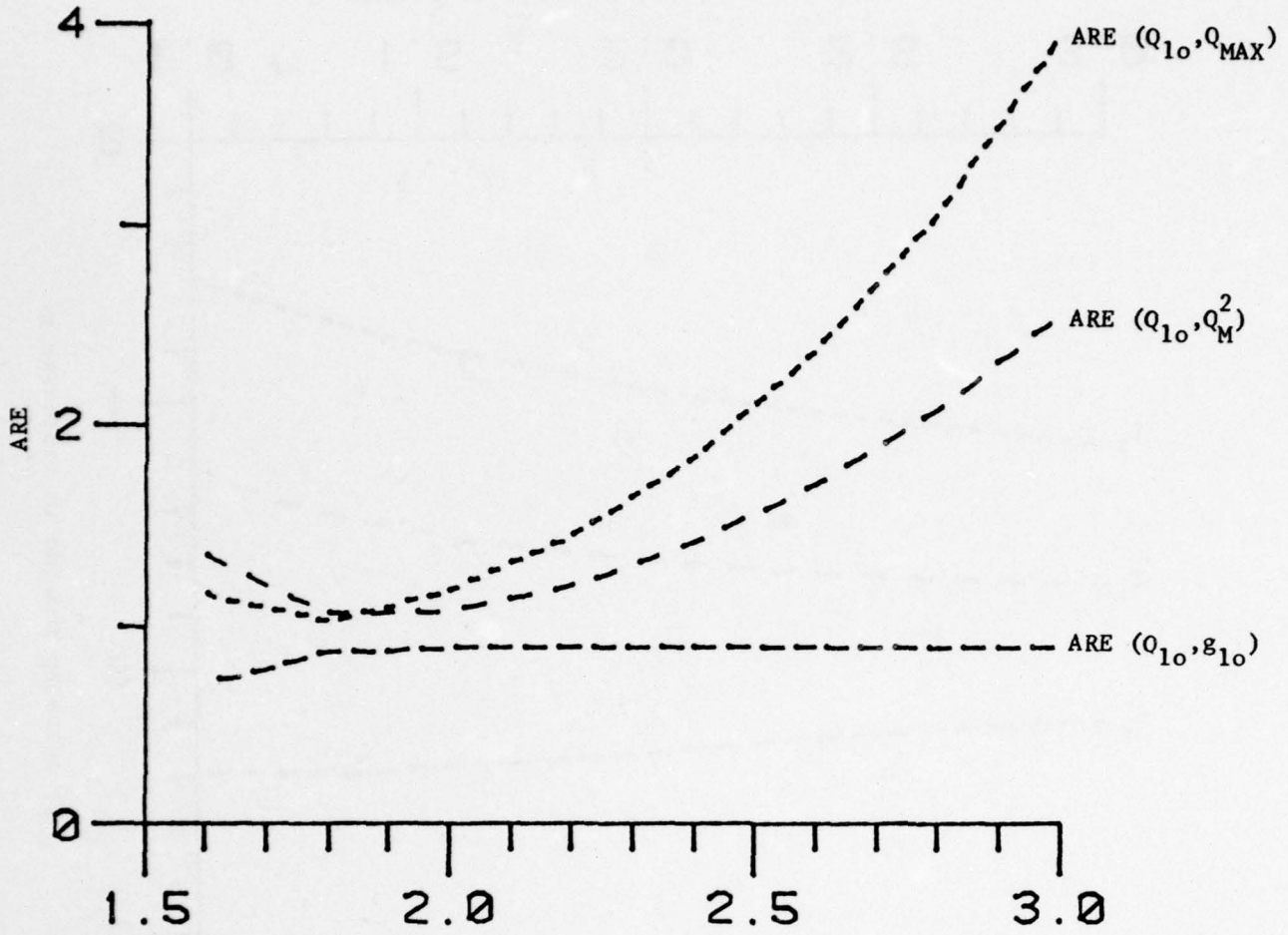


Figure 3. Asymptotic performance comparison.
Stochastic signal, additive noise model.

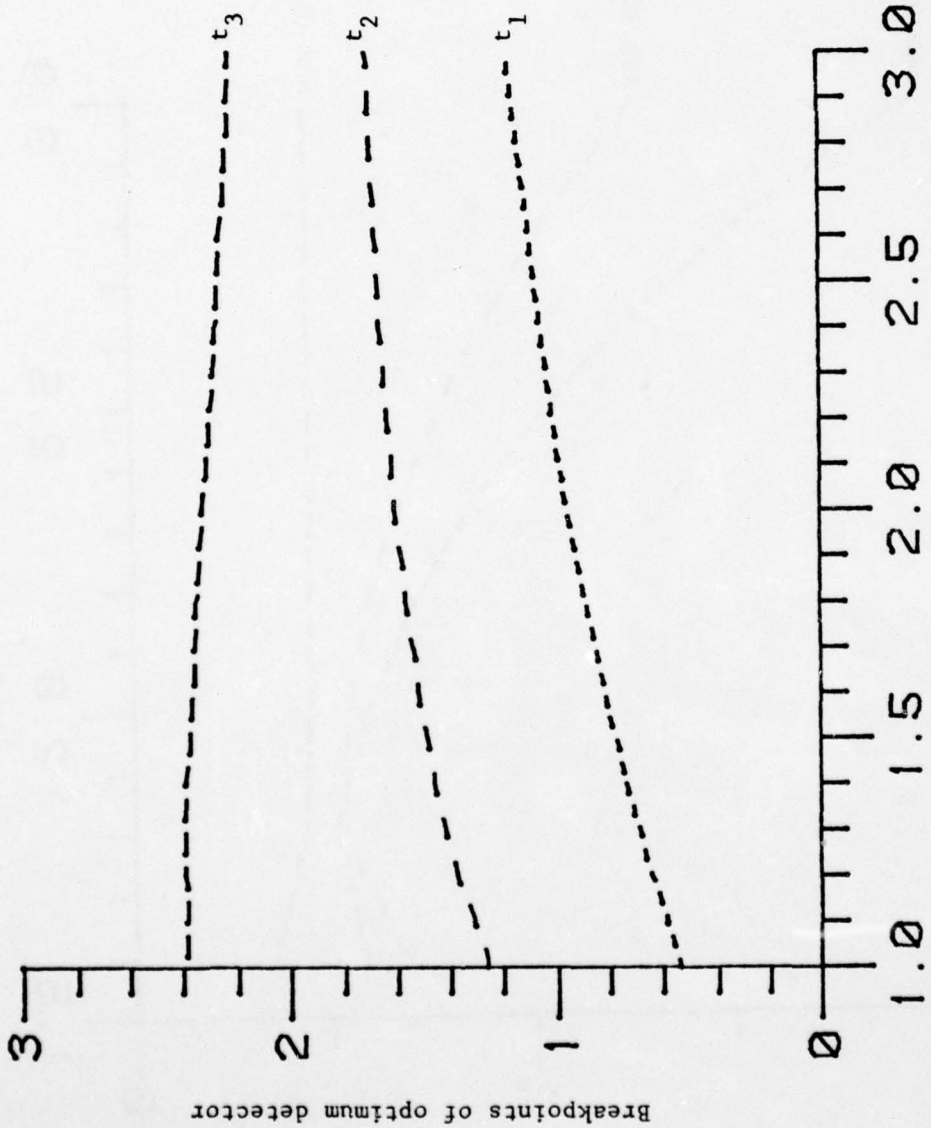


Figure 4. Breakpoints of 4-level locally-optimum quantizer. Stochastic signal, scale-change model.

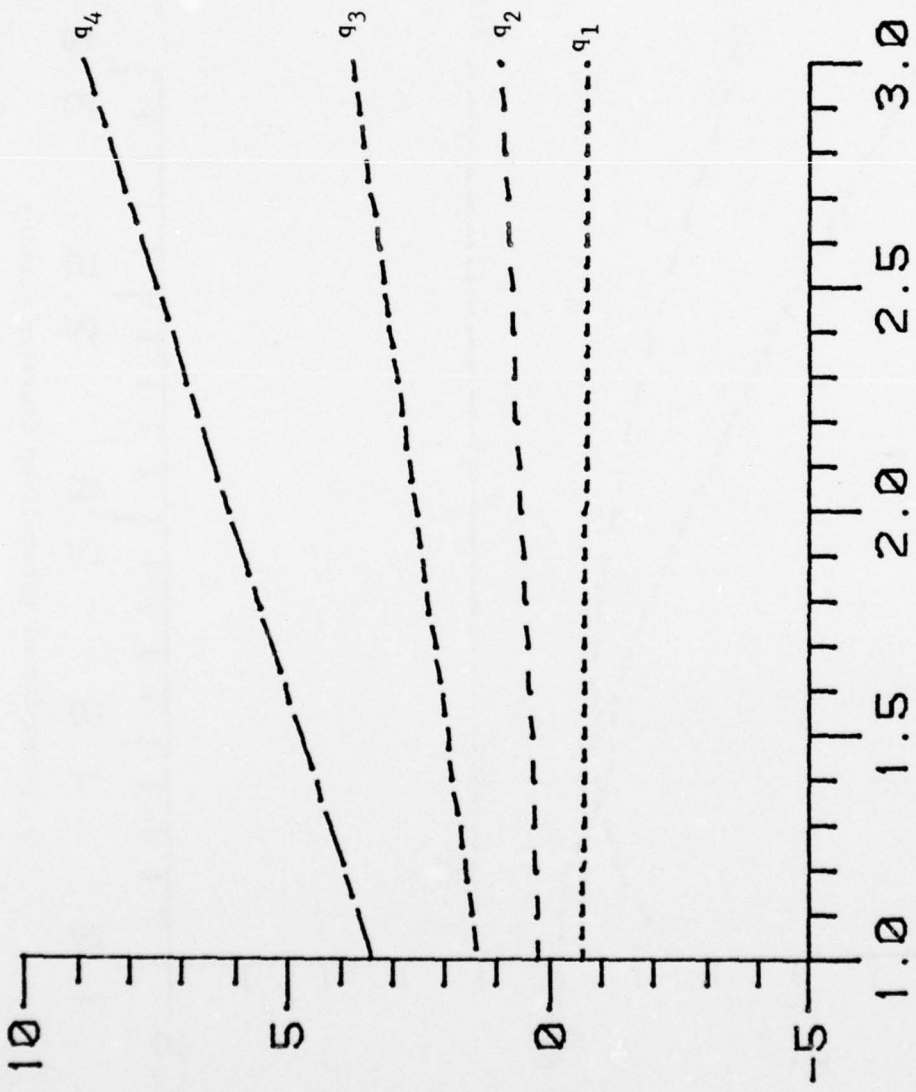


Figure 5. Levels of 4-level locally-optimum quantizer. Stochastic signal, scale-change model.

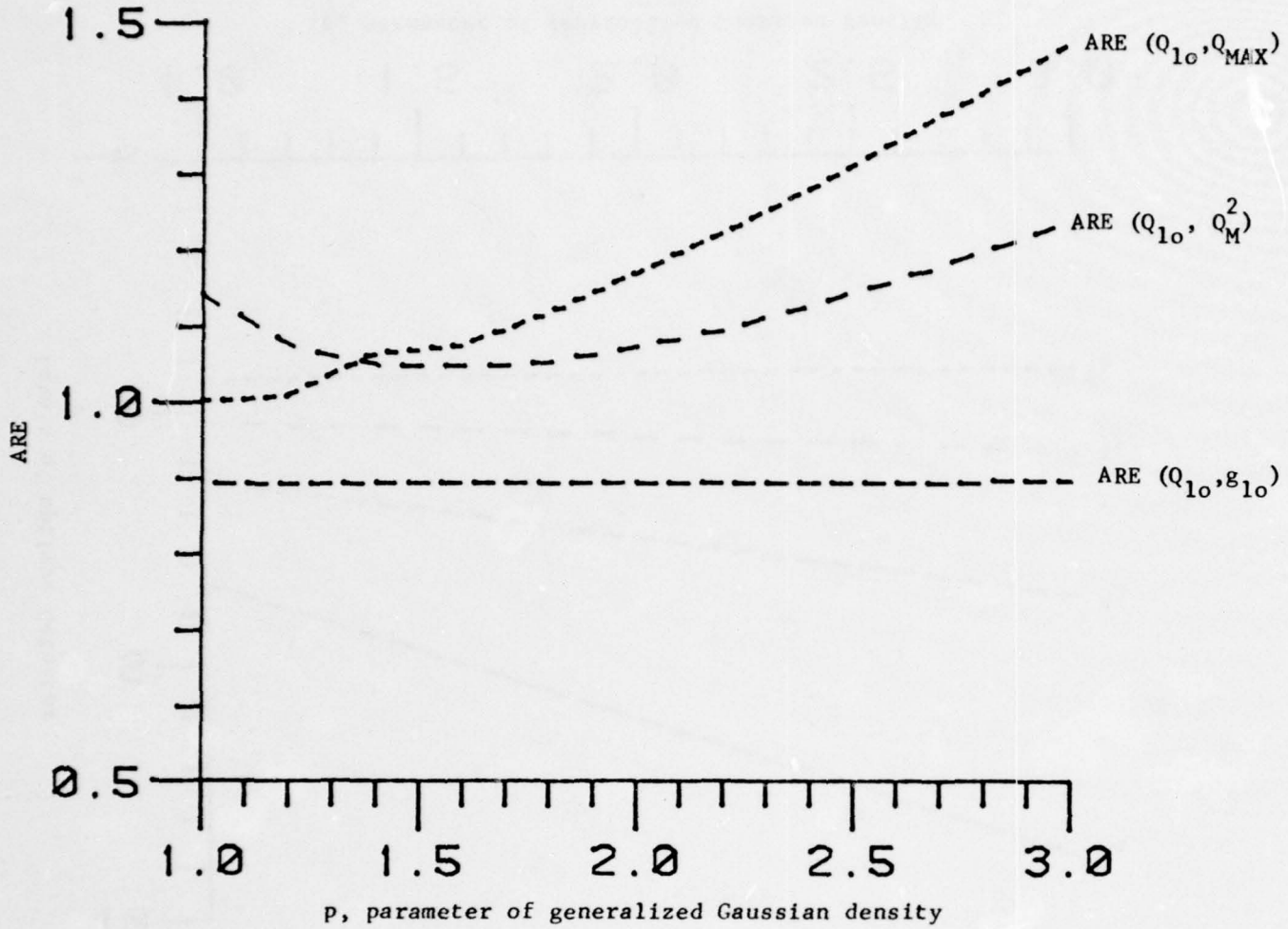


Figure 6. Asymptotic performance comparison.
Stochastic signal, scale-change model.

It can be seen that the parameters of Q_{1_0} for the additive noise case (Fig. 1,2) are more sensitive to variations of the noise density than the parameters of Q_{1_0} for the scale-change model (Fig. 4,5). It is also interesting to note that Q_M^2 performs consistently better (in terms of ARE) than Q_{MAX} (Fig. 3,6). The asymptotic relative efficiency (ARE) between the detectors based on the locally-optimum quantizer (Q_{1_0}) and the locally-optimum nonlinearity (g_{1_0}) respectively, is nearly constant over p and approximately the same for the two cases of interest.

In general, the results indicate that the locally-optimum quantizer produces high efficacy as well as high distortion while the opposite is true for the MMSE quantizer (Q_{MAX}). Keeping in mind the design criteria used for the two quantizers, these results are intuitively satisfying.

5. FURTHER WORK ON QUANTIZATION FOR OPTIMUM DETECTION

In a relatively short period of time after their initial utilization for detection purposes, quantizers have been used to approach several different detection problems.

One such problem, discussed previously here, is the one of robust detection. A "robust" detector, in general, is designed to perform well within a small neighborhood of a nominal model of the noise statistics. The parameters of an optimum quantizer will, of course, also depend on the noise statistics. It seems logical to assume that a quantizer will be robust in some sense because of its inherent insensitivity, at least for a small number of levels. However, a quantizer specifically designed for robust detection would be, in a sense, optimally robust.

Poor and Thomas [24] designed such a quantizer for a class of noise densities. Although they did not show that the obtained robust quantizer is unique, or the best solution available, they showed that it can perform much better than standard detectors in the case of highly contaminated Gaussian noise.

The same authors [25] used quantization to approach the problem of optimum detection in the presence of m-dependent noise.* Note that, in general, an optimal detection procedure for this situation will require a memory of length m. Such systems are not easy to implement, except for spherically invariant noise processes such as the Gaussian. Therefore, it is of interest to derive the optimum detector for this situation from among all memoryless detectors. Based on their earlier work [26] involving the design of general (unquantized) memoryless detectors, Poor and Thomas considered the following hypothesis testing problem for a sequence $\{X_i\}_{i=1}^{n}$

*Note that, here, m denotes the dependence parameter.

of real observations

$$H_0 : X_i = N_i \; ; \; i=1,\ldots,n$$

vs

$$H_1 : X_i = N_i+\theta \; ; \; i=1,\ldots,n$$

where $\{N_i\}_{i=1}^{\infty}$ is a zero-mean, second-order-stationary μ-dependent noise process and θ is a known, positive, constant signal. For this case the following class of detectors was considered:

$$\varphi(Q;\underline{X}) = \begin{cases} 1 \; ; & > \tau \\ \gamma \; ; & \sum_{i=1}^{n} Q(X_i) = \tau \\ 0 \; ; & < \tau \end{cases}$$

where Q is an M-level quantizer. Note that m=0 (independent noise) leads to the problem considered by Kassam. By following the same pattern of thought as in previous cases reviewed here (that is by first assuming fixed breakpoints and by considering the efficacy expression) the authors derived two conditions for the parameters of the optimum quantizer, analogous to the ones derived for the independent noise case. Actually, with the assumption m=0, the two conditions reduce to exactly the same conditions derived by Kassam. Finally, the authors use two examples (stationary Gaussian and Cauchy noise) to show that the m-optimum quantizer performs better than the m=0 quantizer (in terms of ARE) and increasingly so with increasing m.

In most cases discussed here, the performance of decision tests based on quantization was studied on an asymptotic, small signal basis using ARE as a measure of performance. The general quantization problem for binary decisions in the nonasymptotic case was also examined by Poor and Thomas [27]. The main difficulty of this case arises from the fact that probability of error (the natural performance criterion) does not lead to tractable

design procedures. The authors chose to use members of the Ali-Silvey class of distance measures as criteria of optimum detection for the non-asymptotic case. They established necessary conditions for an optimal quantizer design using the criterion of maximum distance as a measure of performance. It was shown that the optimum quantizer for the local case is independent of the choice of the distance measure. However, for the nonlocal case, no single "best" design arises. Nevertheless, these techniques offer a design procedure that can be solved by standard optimization methods.

From these examples, one can conclude that quantizers can be very effective when used with optimum signal selection in mind. Their implementation leads to practical solutions to detection problems for which standard detector design procedures fail in the face of analytical difficulties.

6. CONCLUSIONS

It has been shown that the criteria of maximum efficacy and MMSE between quantized data and data transformed by the locally-optimum nonlinearity are completely equivalent for the general problem of local decisions. A condition for the sufficiency of the (necessary) equations that must be satisfied by the parameters of the optimum quantizer was also derived. Finally, it was shown that for generalized Gaussian noise densities and for the case of stochastic signals (additive noise and scale-change models) a significant difference exists between the detection performance of a locally-optimum quantizer and a minimum-distortion quantizer.

7. APPENDIX

A. Elements of Matrix of Second Derivatives for the Known Signal - Additive Noise Case.

The matrix of second partials, H, is a symmetric triagonal band matrix; i.e.

$$\frac{\partial \eta(\underline{t})}{\partial t_k \partial t_j} = h_{kj} = h_{jk} = \begin{cases} b_k \ , & \text{if } j=k+1 \\ a_k \ , & \text{if } j=k \\ 0 \ , & \text{if } |j-k| > 1. \end{cases}$$

We have calculated the following values for a_k and b_k for Kassam's case.

$$\frac{\partial \eta(\underline{t})}{\partial t_k^2} = a_k = 8[q_k f(t_k)+f'(t_k)]\left[\frac{f''(t_k)}{f(t_k)} - \left(\frac{f'(t_k)}{f(t_k)}\right)^2\right] +$$

$$+ \frac{4[q_k f(t_k)+f'(t_k)]^2 \cdot [F(t_{k-1})-F(t_{k+1})]}{[F(t_{k-1})-F(t_k)]\cdot[F(t_k)-F(t_{k+1})]}$$

$$\frac{\partial \eta(\underline{t})}{\partial t_k \partial t_{k+1}} = b_k = \frac{4[q_k f(t_k)+f'(t_k)]}{[F(t_k)-F(t_{k+1})]}\left[f'(t_{k+1})-q_k f(t_{k+1})-2f'(t_k)\frac{f(t_{k+1})}{f(t_k)}\right].$$

B. Evaluation of Integrals of the Form

$$\int_a^b x^n f_p(x)dx \ ; \ n=0,1,2,\ldots \ ; \ b>a\geq 0$$

where $f_p(x)$ is the generalized Gaussian density of unit variance ($\sigma^2=1$).

Recall that the above density is given by (for $x>0$)

$$f_p(x) = \frac{p}{2\Gamma(1/p)A(p)} \exp\{-(x/A(p))^p\} \ , \ p>0$$

where

$$A(p)=[\Gamma(1/p)/\Gamma(3/p)]^{\frac{1}{2}}.$$

The integral to be evaluated is the following

$$\frac{p}{2\Gamma(1/p)A(p)} \int_a^b x^n e^{-(x/A(p))^p} \, dx \ .$$

Consider the following change of variables:

$$z = (x/A(p))^p$$

$$\Rightarrow \quad x = A(p)z^{1/p}$$

$$\Rightarrow \quad dx = (A(p)/p)z^{(1/p-1)}dz.$$

The integral of interest then becomes

$$\frac{p}{2\Gamma(1/p)A(p)} \cdot \frac{A(p)}{p} \cdot (A(p))^n \int_{(a/A(p))^p}^{(b/A(p))^p} z^{n/p} \cdot z^{(1/p-1)} e^{-z} dz$$

$$= \frac{(A(p))^n}{2\Gamma(1/p)} \int_{(a/A(p))^p}^{(b/A(p))^p} z^{(n/p-1)} e^{-z} dz$$

$$= \frac{(A(p))^n \cdot \Gamma(n/p)}{2\Gamma(1/p)} [\Gamma(n/p; (b/A(p))^p) -$$

$$-\Gamma(n/p; (a/A(p))^p)]$$

where $\Gamma(\cdot)$ is the gamma function

and $\Gamma(\cdot;\cdot)$ is the incomplete gamma function.

8. REFERENCES

[1] J. Max, "Quantizing for minimum distortion," I.R.E. Trans. Information Theory, vol. IT-6, pp. 7-12, March 1960.

[2] G.M. Roe, "Quantizing for minimum distortion," IEEE Trans. Information Theory, vol. IT-10, pp. 384-385, Oct. 1964.

[3] V.R. Algasi, "Useful approximations to optimum quantization," IEEE Trans. Commun. Technology, vol. COM-14, pp. 297-301, June 1966.

[4] a) H. Gish and J.N. Pierce, "Asymptotically efficient quantizing," IEEE Trans. Information Theory, vol. IT-14, pp. 676-683, Sept. 1968.
b) R.C. Wood, "On Optimum Quantization," IEEE Trans. Information Theory, vol. IT-15, pp.248-252, March 1969.

[5] D.G. Messerchmitt, "Quantizing for maximum output entropy," IEEE Trans. Information Theory, vol. IT-17, p. 612, Sept. 1971.

[6] S.S. Rappaport and L. Kurtz, "An Optimal Nonlinear Detector for Digital Data Transmission Through Non-Gaussian Channels," IEEE Trans. Communication Technology, vol. COM-14, No. 3, pp. 266-274, June 1966.

[7] D. Middleton, "Introduction to Statistical Communication Theory," New York:McGraw-Hill, 1960.

[8] P. Rudnick, "Likelihood detection of small signals in stationary noise," J. Applied Physics, vol. 32, pp. 140-143, 1961.

[9] V.R. Algasi and M.Lerner, "Binary detection in white Gaussian noise," MIT Lincoln Lab. Rep. No. 2138, 1964.

[10] O.Y. Antonov, "Optimum detection of signals in non-Gaussian noise," Radio Eng. Electron. Phys. (USSR), vol. 12, pp. 541-548, 1967.

[11] A.K. Ribin, "Classification of weak signals in non-Gaussian noise," Eng. Cybern. (USSR), vol. 10, pp. 901-901, 1971.

[12] T.S. Ferguson, "Mathematical Statistics: A Decision Theoretic Approach," New York: Academic Press, 1967, pp. 235-236.

[13] J. Capon, "On the asymptotic efficiency of locally optimum detectors," IRE Trans. Inform. Theory, vol. IT-7, pp. 67-71, April 1961.

[14] G. E. Noether, "On a Theorem of Pitman," Ann. Math. Stat., vol. 26, 1955, pp. 64-68.

[15] J. H. Miller and J.B. Thomas, "Detectors for Discrete-Time Signals in Non-Gaussian Noise," IEEE Trans. Information Theory, vol. IT-18, No. 2, March 1972.

[16] J. W. Carlyle, "Nonparametric methods in detection theory," in "Communication Theory", A. V. Balakrishnan, Ed. New York: McGraw-Hill, 1968, ch. 8.

[17] Y.C. Ching and L. Kurtz, "Nonparametric Detectors Based on m-Interval Partitioning," IEEE Trans. Inform. Theory, vol. IT-18, No. 2, March 1972.

[18] S.A. Kassam and J.B. Thomas, "Generalizations of the Sign Detector Based on Conditional Tests," IEEE Trans. on Communications, vol. COM-24, No. 5, May 1976.

[19] S.A. Kassam and J.B. Thomas, "Asymptotically robust detection of a known signal in contaminated non-Gaussian noise," IEEE Trans. Information Theory, vol. IT-22, pp. 22-26, Jan. 1976.

[20] S.A. Kassam, "Optimum quantization for signal detection," IEEE Trans. Communications, vol. COM-25, No. 5, pp. 479-484, May 1977.

[21] H.V. Poor and J.B. Thomas, "Optimum quantization for local decisions based on independent samples," J. Franklin Inst., 303, pp. 549-561, 1977.

[22] P.E. Fleischer, "Sufficient Conditions for Achieving Minimum Distortion in a Quantizer," IEEE International Conv. Rec., 1964, pp. 104-111.

[23] D.G. Luenberger, "Introduction to Linear and Nonlinear Programming," Addison-Wesley 1973, pp. 194-197.

[24] H.V. Poor and J.B. Thomas, "Asymptotically Robust Quantization for Detection," IEEE Trans. Inform. Th., vol. IT-24, No. 2, March 1978.

[25] H.V. Poor and J.B. Thomas, "Optimum Quantization for Memoryless Detection in m-dependent noise," Johns Hopkins Conf. Inform. Sciences and Systems, March 1978.

[26] H.V. Poor and J.B. Thomas, "Asymptotically optimum zero-memory detectors for m-dependent noise processes," Proc. 1977 Johns Hopkins Conf. on Inform. Sciences and Systems, pp. 134-139, 1977.

[27] H.V. Poor and J.B. Thomas, "Applications of Ali-Silvey Distance Measures in the Design of Generalized Quantizers for Binary Decision Systems," IEEE Trans. Communications, vol. COM-25, No. 9, September 1977.

[28] H.V. Poor and J.B. Thomas, "Locally optimum detection of discrete-time stochastic signals in non-Gaussian noise," J. Acoust. Soc. Am. 63(1), pp. 75-80, Jan. 1978.